

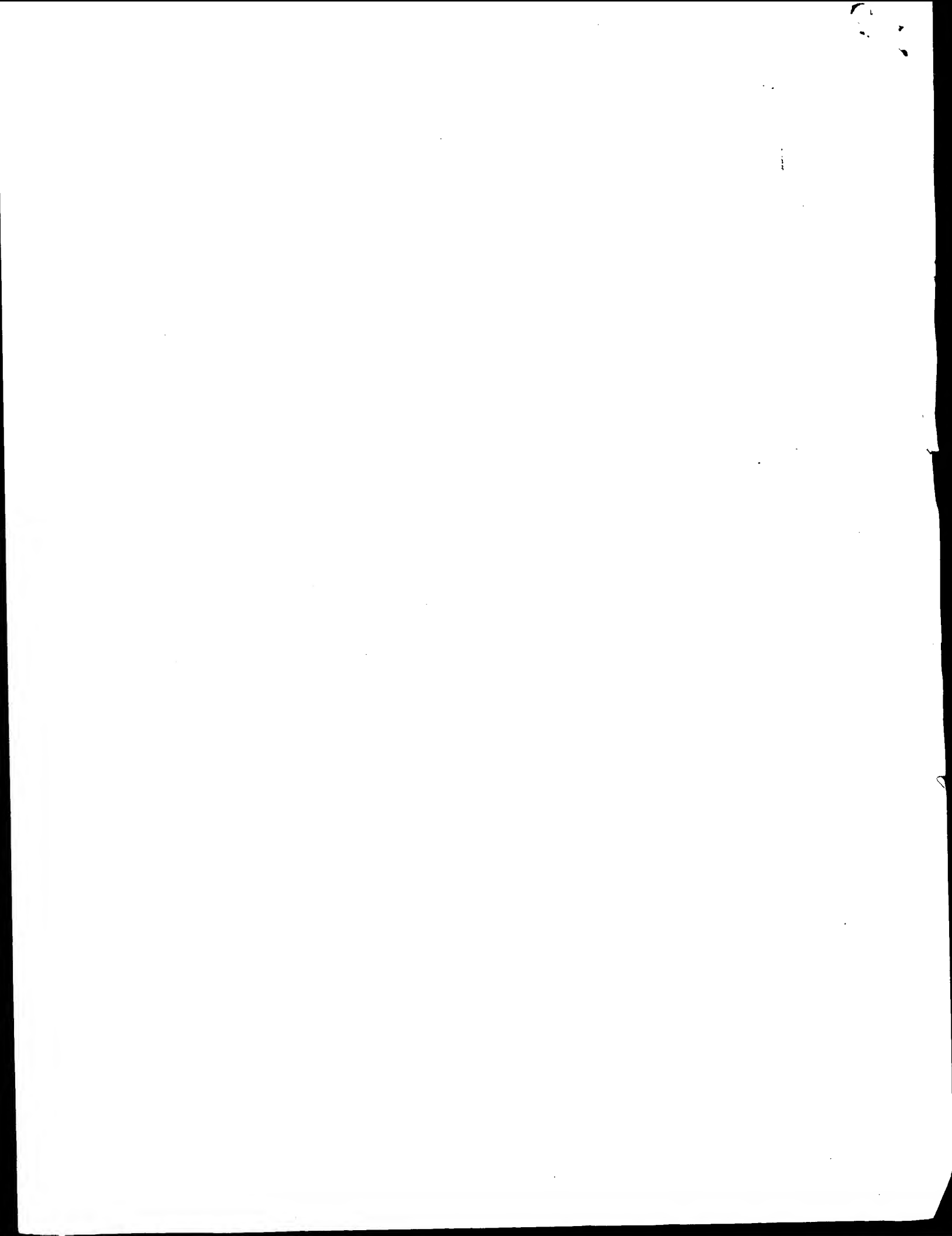
From: Bugaisky, Gabriele  
Sent: Monday, June 04, 2001 10:35 AM  
To: STIC-Biotech/ChemLib  
Subject: 09/596784

These are two microbial proteins that are co-expressed from an operon: please search SEQ ID NO:2 and 4, and interference files of both

thanks, gabi

*Gabriele E. Bugaisky*

- au 1653
- cm1-10d09
- 308-4201



GenCore version 4.5  
 Copyright (c) 1993 - 2000 CompuGen Ltd.  
 OM protein - protein search, using sw model  
 Run on: June 5, 2001, 18:16:39 ; Search time 89.87 Seconds  
 (without alignments)  
 2397.109 Million cell updates/sec

Title: US-09-596-784-2  
 Perfect score: 9448  
 Sequence: 1 MELKSLGTEHKAHVHTAAHN.....NPQVASALTDLKREGLMK 1838

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

#### Database :

SPTREMBL15:\*  
 1: sp\_archaea:\*  
 2: sp\_bacteria:\*  
 3: sp\_fungi:\*  
 4: sp\_human:\*  
 5: sp\_invertebrate:\*  
 6: sp\_mammal:\*  
 7: sp\_mhc:\*  
 8: sp\_organelle:\*  
 9: sp\_phase:\*  
 10: sp\_plant:\*  
 11: sp\_rodent:\*  
 12: sp\_unclassified:\*  
 13: sp\_vertebrate:\*  
 14: sp\_virus:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9448	100.0	1838	2 054581	054581 erwinia amy
2	5529.5	58.5	1829	2 09KH44	09KH44 erwinia her
3	1824	19.3	1795	2 066101	066101 pseudomonas
4	283	3.0	57	2 09LAW1	09LAW1 erwinia amy
5	275	2.9	2514	2 09UJ30	09UJ30 neisseria m
6	255.5	2.7	3455	2 09P906	09P906 xylella fas
7	247	2.6	3442	2 09PBE8	09PBE8 xylella fas
8	245	2.6	2703	2 09K0T0	09K0T0 neisseria m
9	244.5	2.6	3259	4 014789	014789 homo sapien
10	240.5	2.5	5327	5 076891	076891 drosophila
11	240.5	2.5	5476	5 09NJ17	09NJ17 drosophila
12	240.5	2.5	5533	5 09U6C3	09U6C3 drosophila
13	240.5	2.5	5554	5 09NHN1	09NHN1 drosophila
14	237.5	2.5	4152	2 09ZHL3	09ZHL3 haemophilus
15	236.5	2.5	3012	5 097205	097205 leishmania
16	229.5	2.4	2273	2 031152	031152 neisseria m
17	229	2.4	3381	2 09KX33	09KX33 streptococ
18	222	2.3	2540	4 09NQZ2	09NQZ2 homo sapien
19	222	2.3	4823	13 093321	093321 fuju rubrip

20	220.5	2.3	2253	13	P70012	P70012 xenopus lae
21	220	2.3	494	2	Q52389	Q52389 pseudomonas
22	220	2.3	4498	13	O93291	O93291 fuju rubrip
23	219.5	2.3	2614	5	O97054	O97054 dictyosteli
24	218.5	2.3	4957	4	O14687	O14687 homo sapien
25	218.5	2.3	5262	4	O14686	O14686 homo sapien
26	217	2.3	2712	10	O9SB74	O9SB74 arabidopsis
27	216.5	2.3	3634	2	O9JP78	O9JP78 bordetella
28	214	2.3	3282	2	O9PEY9	O9PEY9 xylella fas
29	212.5	2.2	1557	2	O9RNI2	O9RNI2 haemophilus
30	212.5	2.2	2478	2	O9LCH2	O9LCH2 staphylococ
31	212	2.2	2324	11	O9WUJ3	O9WUJ3 rattus norv
32	211	2.2	1794	9	O9TIA7	O9TIA7 bacterioph
33	211	2.2	2647	5	O9U4X0	O9U4X0 plasmodium
34	210.5	2.2	2478	2	O9RL69	O9RL69 staphylococ
35	210	2.2	2059	2	O9PD50	O9PD50 xylella fas
36	210	2.2	2151	5	O9NG79	O9NG79 trichomonas
37	209.5	2.2	1510	5	O61802	O61802 caenorhabdi
38	209.5	2.2	2178	2	O9KWR3	O9KWR3 streptococ
39	209.5	2.2	4919	2	O9ZHL0	O9ZHL0 haemophilus
40	207.5	2.2	3590	2	O45365	O45365 bordetella
41	206.5	2.2	2406	5	O9VXM5	O9VXM5 drosophila
42	206	2.2	3257	5	O9V736	O9V736 drosophila
43	205.5	2.2	4545	2	O9X4W2	O9X4W2 vibrio chol
44	204.5	2.2	2110	5	O9VRA6	O9VRA6 drosophila
45	204.5	2.2	4558	2	O9KS12	O9KS12 vibrio chol

#### ALIGNMENTS

RESULT 1  
 054581 ID O54581 PRELIMINARY; PRT; 1838 AA.  
 AC O54581;  
 DT 01-JUN-1998 (TrEMBLrel. 06, Created)  
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)  
 DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)  
 DE DSPE.  
 GN DSPE OR DSPA.  
 OS Erwinia amylovora.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Erwinia.  
 OX NCBI\_TaxID=552;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=EA321;  
 RX MEDLINE=98115919; PubMed=9448330;  
 RA Bogdanove A.J., Kim J.F., Wei Z., Kolchinsky P., Charkowski A.O.,  
 RA Conlin A.K., Collmer A., Beer S.V.;  
 RA "Homology and functional similarity of an hrp-linked pathogenicity  
 locus, dspE, of Erwinia amylovora and the avirulence locus avrE of  
 Pseudomonas syringae pathovar tomato.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 95:1325-1330(1998).  
 [2]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=EA321;  
 RA Bogdanove A.J., Kim J.F., Wei Z.-M., Kolchinsky P., Beer S.V.;  
 RA Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.  
 [3]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=CFBP1430;  
 RX MEDLINE=98086111; PubMed=9426142;  
 RA Gaudriault S., Malandrin L., Paulin J.P., Barny M.A.;  
 RT "dspA, an essential pathogenicity factor of Erwinia amylovora showing  
 homology with AvrE of Pseudomonas syringae, is secreted via the Hrp  
 secretion pathway in a DspB-dependent way.";  
 RL Mol. Microbiol. 26:1057-1069(1997).  
 [4]  
 RN SEQUENCE OF 1-238 FROM N.A.  
 RC STRAIN=EA321, ATCC 49947;  
 RX MEDLINE=98422475; PubMed=9748455;  
 RA Kim J.F., Beer S.V.;

RT "HrpW of Erwinia amylovora, a new harpin that contains a domain  
 homologous to pectate lyases of a distinct class."  
 RL J. Bacteriol. 180:5203-5210(1998).  
 DR EMBL; U97504; AAC04850.1; -  
 DR EMBL; Y13831; CAA74156.1; -  
 DR EMBL; U94513; AAC62315.1; -  
 SQ SEQUENCE 1838 AA; 198243 MW; 7D595FF78130E8FD9 CRC64;

Query Match	100.0%; Score 9448; DB 2; Length 1838;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 1838; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MELSLGTEHRAAVTAHNPVGHGVALQOQSSSSSPQNAASLAABGKNGKMPRIHQ 60	
DB 1 MELSLGTEHRAAVTAHNPVGHGVALQOQSSSSSPQNAASLAABGKNGKMPRIHQ 60	
QY 61 STAADGISAHHQOKSFSLRGCLGKTKFSRSPQGOPTGTHSKGATLRLDLARDGGETQ 120	
DB 61 STAADGISAHHQOKSFSLRGCLGKTKFSRSPQGOPTGTHSKGATLRLDLARDGGETQ 120	
QY 121 EAAAPDAARLTRSGGVKRRNDDMAGRPVMVKGSGEDKVTQQRHQLNNFQGMQRTMLS 180	
DB 121 EAAAPDAARLTRSGGVKRRNDDMAGRPVMVKGSGEDKVTQQRHQLNNFQGMQRTMLS 180	
QY 181 KWAHPASANAGDRLOHSPPHIPGSHHEIKEEPPVGSTSKATTAHADRVETIAQEDDDSEFQ 240	
DB 181 KWAHPASANAGDRLOHSPPHIPGSHHEIKEEPPVGSTSKATTAHADRVETIAQEDDDSEFQ 240	
QY 241 LHQORLARENPPOPKLGVATPISAREPKLTAVAESVLEGTDTQSPKLPQSMUKGS 300	
DB 241 LHQORLARENPPOPKLGVATPISAREPKLTAVAESVLEGTDTQSPKLPQSMUKGS 300	
QY 301 GAGVTPLAATLDKGLQAPNPPALNTLLKQTLGKDTQHYLAHASSDGSQHLLLDNKG 360	
DB 301 GAGVTPLAATLDKGLQAPNPPALNTLLKQTLGKDTQHYLAHASSDGSQHLLLDNKG 360	
QY 361 HLFDIKSTATSYSVLHNSHPGIEKGLAQAGTGSVSDGSKGISLGGTQSHNKTMLSQ 420	
DB 361 HLFDIKSTATSYSVLHNSHPGIEKGLAQAGTGSVSDGSKGISLGGTQSHNKTMLSQ 420	
QY 421 PGEAHRSLTGTWHPAGAAPOGESIRLHDDKIHLHPGLVWQADKDTSHOLSROAD 480	
DB 421 PGEAHRSLTGTWHPAGAAPOGESIRLHDDKIHLHPGLVWQADKDTSHOLSROAD 480	
QY 481 GKLYALKDNRTLQNLSDNKSSEKLVDKIKSYSDVQRQVAILTDTTPGRHKMSIMPSLDAS 540	
DB 481 GKLYALKDNRTLQNLSDNKSSEKLVDKIKSYSDVQRQVAILTDTTPGRHKMSIMPSLDAS 540	
QY 541 PESHISLSLHFADAHQGLLHGKSELEAQSVAISHGRVADVADSEKLFSAAPKQGDGNE 600	
DB 541 PESHISLSLHFADAHQGLLHGKSELEAQSVAISHGRVADVADSEKLFSAAPKQGDGNE 600	
QY 601 KMKAMPQHALDEHFGHDHQISGFFHDDHQLNALVKNNFRQOHACPLGNDHQFHPGNLIT 660	
DB 601 KMKAMPQHALDEHFGHDHQISGFFHDDHQLNALVKNNFRQOHACPLGNDHQFHPGNLIT 660	
QY 661 DALVIDNOLGLHHTNPPEHIELDMLHGLSLAQEGKLHYFDQLTKGWTGAESDCQLKKG 720	
DB 661 DALVIDNOLGLHHTNPPEHIELDMLHGLSLAQEGKLHYFDQLTKGWTGAESDCQLKKG 720	
QY 721 LDGAAYLLKDGVEKRLNINQSTSSIKHGTENVFSLPHVRNKPEPGDALQGLNKDDKAQAM 780	
DB 721 LDGAAYLLKDGVEKRLNINQSTSSIKHGTENVFSLPHVRNKPEPGDALQGLNKDDKAQAM 780	
QY 781 AVIGVKNYALTEKGDIRSFQKPGTQQLERPAQTLRSREGISGELKDIHVDHKKONYALT 840	
DB 781 AVIGVKNYALTEKGDIRSFQKPGTQQLERPAQTLRSREGISGELKDIHVDHKKONYALT 840	
QY 841 HEGEVFHOPRAWQNGAESSWHKIALPOSESKLSLDSMSHEHKPIATFEDGSOHLKAG 900	
DB 841 HEGEVFHOPRAWQNGAESSWHKIALPOSESKLSLDSMSHEHKPIATFEDGSOHLKAG 900	

QY 901 GWHAYAAPRGPLAVGTSQSQTVPFNRLMQGVKVIPIGSGLITVKLSAQTGTMGAERKV 960	
DB 901 GWHAYAAPRGPLAVGTSQSQTVPFNRLMQGVKVIPIGSGLITVKLSAQTGTMGAERKV 960	
QY 961 SSKFSERIRAYAFNPMTSTPRPIKNAAYATOHGHWQREGKLPLEYMOGALKOLDAHNVR 1020	
DB 961 SSKFSERIRAYAFNPMTSTPRPIKNAAYATOHGHWQREGKLPLEYMOGALKOLDAHNVR 1020	
QY 1021 HNAPODLQSKLETLDLGHEGAELLNDMKRFRDEQSATRSVTVLGHQGVKLSNGEIN 1080	
DB 1021 HNAPODLQSKLETLDLGHEGAELLNDMKRFRDEQSATRSVTVLGHQGVKLSNGEIN 1080	
QY 1081 SEFKSPGKALVQSFVNNRSGQDLKSLOQAVHATPPSAESKLSQMLGHFVSAGVDMSHQ 1140	
DB 1081 SEFKSPGKALVQSFVNNRSGQDLKSLOQAVHATPPSAESKLSQMLGHFVSAGVDMSHQ 1140	
QY 1141 KGEIPLGRORDPNDKTALTKSRLITDVTITGELHELADKAKLVSDHKPDADQIKOLROOF 1200	
DB 1141 KGEIPLGRORDPNDKTALTKSRLITDVTITGELHELADKAKLVSDHKPDADQIKOLROOF 1200	
QY 1201 DTLREKRYESNPVHYTDMGFTHNKALEANYDAVKAFINAFKKEHGHVNLTRTTLVESQ 1260	
DB 1201 DTLREKRYESNPVHYTDMGFTHNKALEANYDAVKAFINAFKKEHGHVNLTRTTLVESQ 1260	
QY 1261 SAELAKKLNTLLSLDGSMSFSRSYGGGVSTVFVPTLSKKVPVPTPGAGITLDRAYN 1320	
DB 1261 SAELAKKLNTLLSLDGSMSFSRSYGGGVSTVFVPTLSKKVPVPTPGAGITLDRAYN 1320	
QY 1321 LFSFRTSGGLNVSGFGDGGVGNIMVATGHDVMPYMTGKTSAGNASDWLSAKHKISPLD 1380	
DB 1321 LFSFRTSGGLNVSGFGDGGVGNIMVATGHDVMPYMTGKTSAGNASDWLSAKHKISPLD 1380	
QY 1381 RIGAAVSGTLOQTQLNSLKFKLTEDELPGEIHLGHTGLTLPAAELKQIEHQMKOGSKLT 1440	
DB 1381 RIGAAVSGTLOQTQLNSLKFKLTEDELPGEIHLGHTGLTLPAAELKQIEHQMKOGSKLT 1440	
QY 1441 FSVDTSANLDRAGINLNEGSKPNGVTARYSAGLSASANAAGRSRSTTSGQFGSTTS 1500	
DB 1441 FSVDTSANLDRAGINLNEGSKPNGVTARYSAGLSASANAAGRSRSTTSGQFGSTTS 1500	
QY 1501 ASNNRPTFLNGVAGAGANLTAALGVAHSSTHEGKPVGTFPAFTSTNVSAALADNRTSOSI 1560	
DB 1501 ASNNRPTFLNGVAGAGANLTAALGVAHSSTHEGKPVGTFPAFTSTNVSAALADNRTSOSI 1560	
QY 1561 SLELKRAEPTVNDISELTSTLKGKFKDSATTMLAALKELDDAKPABQLHLOQHFSAK 1620	
DB 1561 SLELKRAEPTVNDISELTSTLKGKFKDSATTMLAALKELDDAKPABQLHLOQHFSAK 1620	
QY 1621 DVVGDERYEAARNLKKLVIRQQAADSHSMELGSAHSTTYNNLSRINNNDGIVELLHKHFD 1680	
DB 1621 DVVGDERYEAARNLKKLVIRQQAADSHSMELGSAHSTTYNNLSRINNNDGIVELLHKHFD 1680	
QY 1681 AALPASSAKRLGEMMNDPALKDIIKLOQSTPFSSASVSMELKDGLREOTEKAIILDGKVG 1740	
DB 1681 AALPASSAKRLGEMMNDPALKDIIKLOQSTPFSSASVSMELKDGLREOTEKAIILDGKVG 1740	
QY 1741 REEVGVLFODRNNLRVKSVSVSOSKSEGFNTPALLLGTSNSAAMSERNIGTINFKYG 1800	
DB 1741 REEVGVLFODRNNLRVKSVSVSOSKSEGFNTPALLLGTSNSAAMSERNIGTINFKYG 1800	
QY 1801 ODONTPRRTLEGGIAQANPOVASALTDLKKEGLEMS 1838	
DB 1801 ODONTPRRTLEGGIAQANPOVASALTDLKKEGLEMS 1838	

RESULT 2  
 Q9KH44  
 ID Q9KH44 PRELIMINARY; PRT; 1829 AA.  
 AC Q9KH44;  
 DT 01-OCT-2000 (TremBLrel. 15, Created)  
 DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TremBLrel. 15, Last annotation update)  
 DE DSPE.





```

RESULT 3
O66101 PRELIMINARY; PRT; 1795 AA.
AC ID 066101
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE AVIRULENCE PROTEIN (AVRE).
GN AVRE.
OS Pseudomonas syringae (pv. tomato).
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas
OX NCBI_TaxID=323;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DC3000;
RX MEDLINE=98115919; PubMed=9448330;
RA Bordanove A.J., Kim J.F., Wei Z., Kolchinsky P., Charkowski A.O.,
RA Conlin A.K., Collmer A., Beer S.V.;
RT "Homology and functional similarity of an hrp-linked pathogenicity
RT locus, dsPEF, of Erwinia amylovora and the avirulence locus avrE of
RT Pseudomonas syringae pathovar tomato.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:1325-1330(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=DC3000;
RX MEDLINE=98422476; PubMed=9748455;
RA Charkowski A.O., Alfano J.R., Preston G., Yuan J., He S.Y.,
RA Collmer A.;
RT "The Pseudomonas syringae pv. tomato HrpW protein has domains similar
RT to harpins and pectate lyases and can elicit the plant hypersensitive
RT response and bind to pectate.";
RL J. Bacteriol. 180:5211-5217(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=DC3000;
RX MEDLINE=20243785; PubMed=10781092;
RA Alfano J.R., Charkowski A.O., Deng W.L., Badel J.L.,
RA Petnicki-Ocwieja T., van Dijk K., Collmer A.;
RT "The Pseudomonas syringae Hrp pathogenicity island has a tripartite
RT mosaic structure composed of a cluster of type III secretion genes
RT bounded by exchangeable effector and conserved effector loci that
RT contribute to parasitic fitness and pathogenicity in plants.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:4856-4861(2000).
DR EMBL; U97505; AAC06134.1;
DR EMBL; AF232006; AAF71499.1;
SQ SEQUENCE 1795 AA; 195351 MW; 260F74534DE08D5F CRC64;

Query Match 19.3%; Score 1824; DB 2; Length 1795;
Best Local Similarity 29.1%; Pred. No. 5.4e-98;
Matches 554; Conservative 358; Mismatches 722; Indels 272; Gaps 71;

QY 34 SSSPONAASLAAGKNGKMPRIHQSTADGISAHQKKSFSLRGLGTYKFSRSAP 93
DB 61 SKAPOQKAAT-PPTAKN-----VKTPTPPASNVATPRNKARES-----GFSNSSP 103
QY 94 Q-----GPGTTHSGATLRLLARDGDGETOHEAAPDAARLTRSGVKRRND 142
DB 104 QNTHRAPKWLIRNHPNQAASSGAOTHEI-----HPEAAP-----RKNLR 142
QY 143 DMAGRPMVKGSGEDKVPQTKRQHLNFGOMRQMTLSKMAHPASANAGDRIHQSPPHIP 202
DB 143 VREDLPQ-----DRLERSPSYLD 160
QY 203 GSHEHKEEPPVGSSTKATTAHADRVETAQEDDDSEFQQLHOORLARERENPPQP-----PK 258
DB 161 SDNPMTDEEAVANATROF-----RSPDLSHQSGDGTIRISMLATDPPQSSSGSK 209
QY 259 LGVAT-PISARFQPKLTAVAESVLEGTDTTQSPKPSQMLKSGAGVTPLP-AVTLD-KGK 315
DB 1281 NLTPPTFIILADK--ATGLWPTAGATGNRNYILNAERCEGVTLYLISEGA-GNVSQGP-- 1281

```

```

DB 210 IGDSDGPIPPR-----EPMLWRNGGGRFELKDEKLVNRNS-----EPQSGIQLDAK 256
QY 316 LQAPDNPPALNTLLKOTLGKDTQHYLAHASSDGSQSOHLLDNKGHLFDIKSTATSYSL 375
DB 257 PDFSTFNTGCLAPLLDSILATPKQTYLAHQSKDGVHGQLQANGHFLHQAODSSSLAVI 316
QY 376 HNSHPG-EIKGKLAQAAGTSVSDGKSGKISLGSQTQSHNKTMLSQPOGAHRSLLTGIWQ 434
DB 317 RSSNEALLIEGKKPPA---VKMEREDGNIHI--DTASGRKTO-ELPGKAHLAHTNV-- 367
QY 435 HPAGAARPOGESIRLHDDKTHILHPGLVWQSD--KDT-HSOLSRQADGKYALKDNR 491
DB 368 ----LLSHDGERMRVHEDRLYQFPISTRWKIPGEGLEDTAFNSLTSGGNGSVYAKSDAV 423
QY 492 LONLSNKSSEKLVDKTKSYSDQRGQVAILTTPGRHKMSIMPISLDASPESHISLSLHF 551
DB 424 V-DLSFPFPHVEVEDLOFSVAPDNRAALLS---GKTTQAILLT-DMSP-----VIGGLT 474
QY 552 ADAHQGLLHGKSELEASQVAISHGRLVVADSEKGLFSAAIPKQDGNELKMKAMPOHALD 611
DB 475 PKTKGLELDGGRAQAAGVGLSGDKLFTADTQGRLYSAD-RSAFEGDDPKLKLMPQANF 533
QY 612 E----HFCHDQHSFGFHDHGHQNLVKNFRQOHACPLG-NDHQFHPGWNLTDLVID 666
DB 534 QLEGVPLGGHNRVTGFIINGDDGVHALLKRNQGETHSHALDEQSKSQSGWNLTNALVN 593
QY 667 NQLGLHHTNPEP---HEILDMGHLSLALQEGKLHYFDQLTKGWTGAE-SDCKQLKKGLD 722
DB 594 NNRGL--TMPPTTAADRLNLDRAGLVGLSEGRIQWDATPECWKADIKDRLQRGAD 651
QY 723 GAAYLLKDGVEKVRNINQSTSSIKHGTENVFSLPHVRNKPPEGDALQGLNKDKA-QAMA 781
DB 652 SNAYVLKGGKLHALKIAAEHPNMAFDRTALTAQTARSTKVEMGKEIEGL--DORVIAKFA 709
QY 782 VIGWNKYALTEKGDIRSFOIKPGTQQLERPAQTSREGISGELKDIHVDHKNLYALTH 841
DB 710 MVSNNKRFVALDD-----QNKLTASHDKHPVTLDIPLEGDIKSLSDKHNHALTS 762
QY 842 EGEVFHQPREAWN---GAE-SSSWHKLALPOSESKLSLSDMSHEKPIATFEDGSHQ-- 895
DB 763 TGGLYCLPKPAQWSTKLGDLQRLARPTVALPGQ--PVKALFTNDDNVLSAQIEDAGKGL 821
QY 896 -OLKAGWHAYAAPERGPLAVGTSGQVFNRLMQGVKGVPGSLTVKLSAQTGGMGTG 954
DB 822 MOLKAGQWQRF---EQRD--VEENGLNDVHSRITSNKTWRIPKTLTLMQDVTYFGRSG 876
QY 955 AEGKVSXSKSERIRAYAFNPTMSTPRPIKNAAYATQHGQWQREGKLPYEMGALIKQL 1014
DB 877 VERSKKAST-SEFIRANIYKNTAETPRWMKNVGDHIQHRYOGRGLKEVYETESMLFKQL 935
QY 1015 DAHNVNRINAPOP-----DLQSKLETLD--LGEHGAELNDMKRFRDELEQASATSVTLG 1067
DB 936 EL--IHESGRRPPARGQDLKARITALEAKLPGQATLVKELETRDELENHSTALMSIG 993
QY 1068 -----HQGVKLSNGSEINSEFKPSP-----GKALVQ---SFVNNRSGQDLSKLSLQ 1110
DB 994 QSYKAKNLKQDGLINOHGEL---AKFSVRMQFGKKLADLGTKLKFKSSGHDVLKELQD 1050
QY 1111 AVHATPPSAESKQSLQMLGHFVSAGVDMSHQKGIPLGRORDPNDKLTALTKSRLLDTVTI 1170
DB 1051 ALTQVAPSAENPTKLLGLTKHQLKSHQKADIPLQRRDASEDHGLSKARLALDLVTL 1110
QY 1171 GELHELADKALVSDHKPDADQIKOLROQFDTLREKRYENPNVKHYTDMGTFHKNKALEAN 1230
DB 1111 KSLGALLDQV----EQLPPQSDIEPLQKQLATLRDVTYENPVVVTDMGTFDNKALESG 1166
QY 1231 YDAVKAFINAFKKEHHGVNLTTRTVLESQSAELAKKLTNLSLDSG-ESMSFSRSYGG 1289
DB 1167 YESVKTFKSKADHAVSVNNRAATGSKDQAELAGKFKSMLQLEHGDGDEGLQSVGV 1226
QY 1290 GVSTVFVPTLSKKVVPVPIPGACITLDRAYNLISFSTSGGLN---VSFGRDGGVSGNIMV 1346
DB 1227 NLTPPTFIILADK--ATGLWPTAGATGNRNYILNAERCEGVTLYLISEGA-GNVSQGP-- 1281

```

```

QY 1347 ATGCHDVPYM-----TGKKTSAGNASDWLSAKHKISPDLRIGAAVSGTLOGTQNSLKFK 1401
DB 1282 GAGKYWPGFFDANNPARSDVGN-----NRTLTPNFKLGVDVTAIVAASRAGVVFN 1334
QY 1402 LPEDELPFGIHLTHGLTPAELLQKQIEHOMKQGSKLTFSDVTSANLRLRAGINLNE- 1460
DB 1335 VPDEDIDAFVDLFEQOLNPLOVLKKAVDHESYEARRFNFDLTAGGTADTRAGINLTEDR 1394
QY 1461 --GSKPN--GVTVARYSAGLSA--SANLAAGSRERSSTSGFGSTTSANNRPFLNGVGA 1514
DB 1395 DPNADPNSDSFSAVVRGGFAANITVNLMTYTDYSLTKNDKTELKEGKNRPFLNNVTA 1454
QY 1515 GANLTAALGAHVSSTHEGKPV---GTFPA--FTSTNVSAAL--ALDNRSTQSSTLELKRA 1567
DB 1455 GQQLRAQIGGSHAP--TGTPASAPGTPASQTAANNLGGALNFSVENRTVKRKFRYNA 1513
QY 1568 EPTVNDISELSTLGHKFKDSATTMLAALKELDDA-----KPAE-----QLHILOQHS 1618
DB 1514 KPITTEGLSKLKGLEAFDLNWTAKLAELADPLNARYTGKKPDEVIQALDGLLELFA 1573
QY 1619 ---AKDVVGDEREAVRNKLKLVIRQQA--ADSHSMELSGASHSTYNNLSRNNNDGIV- 1672
DB 1574 DTPPKD--NDKQYKALRDLKRAAVERASANKHSV--MONARFETSKINLSGUSSEILT 1630
QY 1673 ELLKHFDALPASSAKRILGEMNNDPALDKIQLQSTPFFSASVSMELKDLREOTEK 1732
DB 1631 KIMSSVROASAP--GNATRVAEFMRQDKLRLAMEGSGTGLTARVLEPKDSLVDKIDE 1689
QY 1733 AILDGKVGREEVGVLFQDRNNLRVKSVSQSVKSEGFNTPALLGTSNSAAMSERNI 1792
DB 1690 GSLNGTMTQSDLSMLEDRNEMRIKLVVFHTATQAEFTSPPLVSYNSGANVSVTKL 1749
QY 1793 GTINFKYGODQNTPRFTLEGGTAQANPOVASALTDLKEGLEMKS 1838
DB 1750 GRINFEVIGADQKPIGYTFDGLSRPSASLKEAAGDLKKEGFELKS 1795

RESULT 4
Q9LAW1
ID Q9LAW1 PRELIMINARY; PRT; 57 AA.
AC Q9LAW1:
DT 01-OCT-2000 (TremBrel. 15, Created)
DT 01-OCT-2000 (TremBrel. 15, Last sequence update)
DE HRP-SECRETED PATHOGENICITY/AVIRULENCE PROTEIN DSPE (FRAGMENT).
GN DSPE.
OS Erwinia amylovora.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Erwinia.
OX NCBI_TaxID=552;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EA246;
RA Kim J.F., Laby S.V.;
RT "Comparison of the hrpN-flanking regions of two Erwinia amylovora
RL strains with different host specificity";
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF083620; AAF63403.1;
FT NON_TER 57
SQ SEQUENCE 57 AA; 5788 MW; AFF0CA36311E4BE1 CRC64;

```

```

Query Match 3.0%; Score 283; DB 2; Length 57;
Best Local Similarity 98.2%; Pred. No. 3.7e-10;
Matches 56; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MELKSLGTEHKAHVTAHNPVGHGVALQOQSSSSPQNAASLAAGKNGKMPRI 57
DB 1 MELKSLGTEHKAHVTAHNPVGHGVALQOQSSSSPQNAASLAAGKNGKMPRI 57

RESULT 5

```

```

Q9JY30
ID Q9JY30 PRELIMINARY; PRT; 2514 AA.
AC Q9JY30:
DT 01-OCT-2000 (TremBrel. 15, Created)
DT 01-OCT-2000 (TremBrel. 15, Last sequence update)
DT 01-OCT-2000 (TremBrel. 15, Last annotation update)
DE HEMAGGLUTININ/HEMOLYSIN-RELATED PROTEIN.
GN NMB1768.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGROUP B;
RC MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwin M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Cittone H., Clark E.B.,
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Maignani V., Pizzi M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RA "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58.";
RT Science 287:1809-1815(2000).
RL EMBL; AE002526; AAF42109.1;
DR TIGR; NMB1768;
SQ SEQUENCE 2514 AA; 265615 MW; 95643A671B3BC268 CRC64;

Query Match 2.9%; Score 275; DB 2; Length 2514;
Best Local Similarity 18.8%; Pred. No. 6e-07;
Matches 389; Conservative 287; Mismatches 803; Indels 586; Gaps 91;

QY 7 GTEH--KAAVHTAAHNPV-----CHGVALQOQSS-----SSPNAAS 43
DB 37 GSAHVKSVPFGTTHAPVCRSNIFSFLLGSLCLAVGTANIAFADGIADKAAPKQAT 96
QY 44 LAAGKNGKMPRIHOPSTAADGISAHQKKSFSRLGCLGKTKKFSRSAPQGP----- 98
DB 97 ILQGTG---IPQVNIQTPTSAGSVNQVQDFVNGRGAAILNN--SRSTQTQLGWIQ 151
QY 99 -----TTHSK-----CATRLDLLARDGGETQHEAANPDAARLTR 132
DB 152 NPWLARGEARVVVNOINSSHSSQMGYIEVGRRRAEVVIANPAGIAVNGGFINASRATL 211
QY 133 SGGKRRNMDMAGRPVKG-----GSGED-----KVPTQOKRHLNFFGMRQTMLS 180
DB 212 TTGQPOVQAGDLGFKIRQGNVVIAGHGLDARDTDFTRLSYHSKIDAPVWGQDVRVAG 271
QY 181 KMAHPASANAGDRLOHSPHIPGSHHEIKEEPPVGSSTKATTAHADRVETAQEDDDSEFOQ 240
DB 272 QNDVVATGNA-----HSP-----ILNNAANNTNNNTANGTHIPLFAIDTG----- 312
QY 241 LHOQRLARENPQPPKLGATPISARFQPKLTAVAESVLEGTDTTQSPKLPQSMKGS 300
DB 313 -----KLG-----GMYANKITLILSTA-----EQAGIRNQGLFAS 342
QY 301 GAGVTPLAVTLDKGLQALAPDNPALNTLLKOTLKGCTOHYLAHHAS--SDGSQHLLDN 358
DB 343 SGN-----VAIDANGRLVNSGTMMAAN-----AKDNTAEHKVNIISQ-----VEN 385
QY 359 KG-----HLFDIKSTATSYS-----VLHNSHPGEIKGLAAGTSVSDGSGKI 404
DB 386 SGTAVSQOQTQIHSQSIQNTQTLSSGEILHNS-----GSLKNETSGTI-----EAARL 435
QY 405 SLGSGTOSHNTKMTLSQPGEAHRSLLTGTWQHPAGAAPQGESIRLHDDKHLHILHPGLY 464
DB 436 AIDTDT--LNNQGLSQTG-----SQKLHI----- 458
QY 465 QSADKDTHSQLSROADGKLYALKDNRTLQNLSD-----NKSSEKLVDKIKSYSDORQ 518

```

Db 459 -----DAQKM-----DNRGRMGLODPTASNGSSNOTGN-----SYNASPHSS 498  
Qy 519 VAILTDTGRHKMSIMPDLSDASPESHISLSLHFADAHQGLLHCKSELEAQSVAISHGRV 578  
Db 499 TTTPTATGTATVSINITAPT-----FAD-----GTIRTHGALDNGSIIANGQTD 547  
Qy 579 VADSEKFLSAIPKOGDGNELKMKAMPQHALDEHFHQHDIISGFFHDDHGGQNALVKN- 637  
Db 548 VSAQGG-LNNAG--QIDHOLNAG--SAPDNHNG--TIISDAVHIQAGSLNNGN 598  
Qy 638 NFRQOACPLGNDHOFHFGWNTDALVID-NOLGLHHTNPEPEHIDMGLGSLALQEGK 696  
Db 599 TTRQO-----LEIETDQDQAHGKLLSAEIALDLAVSGSLNNGE 638  
Qy 697 LHYFDOLTKGWTGAESDCKQKKGGLDGAAYLLKDGVEKRLNINOSTSIKHTENVESLP 756  
Db 639 IATNOOL-----IHDGQOSTAVIDNTNGTIOGSRDVAIOAK 675  
Qy 757 HVRNKPPECDALQGLNK-----DRAQAMAVIGVKNYLALTEKDIRSFQIRPGTQOLE 810  
Db 676 SLSN-----NGTLAADNKLDIALQDDFYVERNIVAGNE-LSLSRGSLSKNSHTLQAGKRIR 730  
Qy 811 RPAQYLSREG-----ISGELKDIHVDHKQNLIALTHEGEVPHQPREAWONGAESSSWHKLA 866  
Db 731 IKANNLDNAAQGNISGGTDTICQHN-----LTNRGLIDQ----- 767  
Qy 867 LPQSESKLSLDMSH-----EHKPIATFEDGSOHLKAGGWHAYAPERGPLAYGTS 918  
Db 768 ----QTKIQAGQMNIGTRIYGDNTAIATRLDNDQDENGTA--ATAARENLMGIG-- 819  
Qy 919 GSQTVFRLMOGVKVIPIGSLTVKLSAQTGMTGAERKVSCKFSERIRAYAFNPTMS 978  
Db 820 ----QLNNRENSLIYSGNDMVGALDNTNGOATCKAQR----- 853  
Qy 979 TPRPKNAAYATOHGWOGRGLKPLYEMOGALIKOL-----DAHNVHNP 1024  
Db 854 ----IHNAGATIEAAGKMLGVEKLHNTNEHLKTQLVETGREHIVDYEAFCRHELLREGT 909  
Qy 1025 QPDL-----QSKLETLDLGEH-----GAELNDMK 1049  
Db 910 QHELGSVYNDESDDLRTPDGAAHENHKKYDEKVTQKTQVOTAPAKIISGNDLTIDGK 969  
Qy 1050 R-FRDEQSATRSVTYVLCQHO-----VLKNGEINSEFKPS-PGKALVOSFN 1096  
Db 970 EVFNTDSQIAGNLIVQTEKGLHNEQTFGKKVFSENGKLSYWEKHKGR---DSTG 1026  
Qy 1097 VNRSQDLSKSLQQAQVATPPSAESLQSMGLHFVSAGVDMSHQKGIPLGRQRPDNDKT 1156  
Db 1027 HSEQNYTLPEETRNISLSGFAYESHKALSHHAPSGQTELPQSG-----ISLPTSNSFT 1083  
Qy 1157 ALTKSRL-ILDTVTITIGELHELADKAK-----LVSDHKPDA---DOIQLRQOQFDTLREKR 1207  
Db 1084 PLPSSLYIINPVNGYLVTDFRANYRQWLGSYDMLDSKLDPNLKRGLGDIYEQR 1143  
Qy 1208 YESNPVKHYTMGFTHINKALEANYDAVKAFINAFKHEHGVNLTTRVLSQGSABEAKK 1267  
Db 1144 LINEQIAELT--GHRRLDGYQNDQEOFKALMDNGATAARSNNLSVGLAL-----SAEQVAQ 1197  
Qy 1268 LKNTLLSLDSGSMSPSRSGYGVSVFVPTLSKKVPVPPVPGAGITLDRAYNLSFESRST 1327  
Db 1198 LTSIDIWLVOKE-----VKLPDGGTQTVLPVQVYVRKNGLDIGKAL-----LSSGNTQ 1247  
Qy 1328 GGLNYSFSGDGGVSGNIMVATGHVMPYMGK-----KTS-----GNASDWLSAKHKIS 1377  
Db 1248 INVSGLKNSGTIAGRNALLINTDLDNIGGRHIAQKSAVTATQDINNIGMLSAEQTL 1307  
Qy 1378 PDLRIGAAV-SGTLOQTQLNSLKFKLTEDLPG-FIHLHGTITL-----PAELLKGIE 1430  
Db 1308 --LNAGNNINOSTTASSQNTQGSSTYLDRAAGIYITGKGVLAQAQKDKDINIAGQIS 1365  
Qy 1431 HOMKQSGKLTFSVDTSANLDL-----RAGINLEDGSKPNCVTVARVAGLSASA----- 1479  
Db 1366 NOSEOG-QTRLAQGRDINLDTVQTSKHQATHFDADNHVIRGSTNEVSSIQTKGDVTL 1424

Qy 1480 --NLAAGSRERSTTSGOFGTTSGASNNRPTFLNCVGAGANLTAALGVVAHSTHEGKPGVI 1537  
Db 1425 GNNLNAKAEEVSSANTL--AVSAKND-----IN-ISAGINTT--HVDASKHTGRSGG- 1473  
Qy 1538 PFAFTSTNYSAALALDNRSTQSISLELKRAEPTVTSNDISELTSTL-----GKHFK 1587  
Db 1474 ---GNKLVTIDRAQSHHETAQSTFEKGQVVVLOAGNDANILGNSVINDSGTQIOAGNHVR 1530  
Qy 1588 DSATTKMLAALKELDDAKPAEQHLILOHFSKADVDVDEREAVRNKLKLVIRQQAASH 1647  
Db 1531 ICTT-----QTOSQSEYHQTKSGLMSAGI-----FTI---GSKTN 1565  
Qy 1648 SMELGSASHSTTYNNLSRINNDDGIVELLHKKHFDALPASSAKRIGEMM-----NNDPALK 1702  
Db 1566 TQENQSQSEHTGSTVGLSKGDTTI-VAGKHYE-----QIGSTVSSPBGNTIYAQ 1615  
Qy 1703 DIKQOSQTPFFSASVSMELKDLGRLQRETERKAILDGKVGREVG-----LPQDRNNLRV 1756  
Db 1616 SIDIQAHNKLNNTTQTYEQKGLTFAVSPVTD--LAQQAIAVAQSSKQVQSGKND-RV 1672  
Qy 1757 KSVSVSOS--VSKSEGFNTPALLGTSNAAAMSMERNIGTINFKYQDQDTPRRT-LRG 1813  
Db 1673 NAWAANAGQAYQTKSAQNLANGTTNAKQVS-----ISITYGEQON--RQTQVOA 1723  
Qy 1814 GIAQANPOVASALTDLKKEGLEMS 1838  
Db 1724 NQAQASQIQAGGKTTLIATGAEOS 1748

RESULT 6  
Q9P9U6 PRELIMINARY; PRT: 3455 AA.  
AC Q9P9U6;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)  
DE HEMAGGLUTININ-LIKE SECRETED PROTEIN.  
GN XF2775.  
OS Xylella fastidiosa.  
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;  
OC Xylella.  
OX NCBI\_TaxID=2371;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=9A5C;  
RX MEDLINE=20365717; PubMed=10910347;  
RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,  
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,  
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carier H.,  
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,  
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorfi H.,  
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,  
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,  
RA Garnier J.S., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,  
RA Ho P.L., Hohelsel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,  
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,  
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,  
RA Machado M.V., Martins E.A.L., Martins H.M.F., Marino C.L.,  
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,  
RA Meen C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,  
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,  
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,  
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,  
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,  
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,  
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,  
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,  
RA da Silva J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,  
RA de Souza A.P., Terenzi M.F., Truffi D., Tsal S.M., Tshako M.H.,  
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,  
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;

\*The genome sequence of the plant pathogen Xylella fastidiosa.\*;

RT Nature 406:151-157(2000).  
 DR EMBL: AE004082; AAF85560.1; -  
 DR INTERPRO: IPR000267; -  
 DR INTERPRO: IPR001424; -  
 DR PROSITE: PS00087; SOD\_CU\_ZN\_1; UNKNOWN\_1.  
 DR PROSITE: PS00144; ASN\_GLN\_ASE\_1; UNKNOWN\_1.  
 SQ SEQUENCE 3455 AA; 360947 MW; 4CD692CB8752FDAA CRC64;

Query Match 2.7%; Score 255.5; DB 2; Length 3455;  
 Best Local Similarity 18.6%; Pred. No. 1.4e-05;  
 Matches 412; Conservative 285; Mismatches 825; Indels 695; Gaps 101;

QY 3 LKSLGTEKHAHVAHNPVGHVVALQOQSSSSPONAAASLAAGKNGKMPRIHOPST 62  
 DB 1374 IDNLGTRLYGDHIAL-----HAQTLNREDETSDGTHAATAIARQRDLICADTLRWAN 1428  
 QY 63 A---ADGISAHHQOKSFSLRGCL---GTRKFF--SRSAPOGQGTTHSKGATLRLDLLAR 114  
 DB 1429 AMILSDGDAAI-----GATLDNALHATGATLTDNRSATIDTGLNITITLNNI--RD 1481  
 QY 115 DGETHEEAAPDA---ARLTRSGVVRNMDMAGRMVWGGGDEKVPQOKRHQLNNF 171  
 DB 1482 N---VHIAHAPDVVTEARMEQPHW--RKNQPN-----GGSGNFRFTSNYDAHDI--- 1525  
 QY 172 GQMRTMLSKMAHPASANAGDRLOHSPPHIPGSHHEIKKEPVGSTK-----A 219  
 DB 1526 -----YYLPADIIKDD-----PYITPDGQQLHRAIVRLTPOTSAYFYARGGLYA 1570  
 QY 220 TTAHADRVETAEQDDDEFQQLHQQLARENPPPKLGVATPISARFPQKLTAVAES 279  
 DB 1571 SOAERRRMDLTARTGDSVLYYYDR-----QDKQPNPDHVA--ATNHS 1613  
 QY 280 VLEGDTTQSPPLKQPSMLKSGAGVTPLATVLDKGLQLAPDNPALNTLLKQTLGKDTQ 339  
 DB 1614 AFIGLDTPOONERFQT-----VPITYAPGDRLTYDSNYG-----TCTDDCV 1655  
 QY 340 HYLAAHSSDGSQHLHLDNKGHLFDIKSTATSYSLHNSHPGEKGLQAAGTGSVSDG 399  
 DB 1656 RLVTWHDYTD-PDHTLID-----MHRG--PNDR-----DN 1693  
 QY 400 KSGKISLGSQTSNKTML-SOPGEAHSRLTGTWHPAGAAPQGESIRLH-----D 451  
 DB 1684 EKYR-----DATRTQDILNPDAGAPALIQTG-----GAMMIQTDTFLRNHYADLLAGGD 1733  
 QY 452 DKIHILHPELVWOSADKTHSOLROADGKLYALKDNRTLQNLSDNKSSEKLVDKLSY 511  
 DB 1734 QTVGLPPLPHTKEKSDEHKY-----KRVLLIDNRALQ--LS-----RTDTFPHI 1776  
 QY 512 SVDQROGA-----ILDTPCGRHKMSIMPSLDASPEHSISLSHFADAHQGLHG 561  
 DB 1777 STTYRGKVSPEWSNESRTPTTQIGRITSGGHQHTAAQTLLNNVTDSTHAPETIQLHTYN 1836  
 QY 562 KSELEAQSAVSHGRVLVADSEKGLFSAAPKPGCDGNEKMKAMPQHALDEHFGHDHQS 621  
 DB 1837 PS---TQTLSSVYGVITVTDTPSLTSLVSLADNPGSAGQELTYIPQSIPTP----- 1885  
 QY 622 GFPHDDHGGOLNALVK-----NN-----FROQHACPLGNDHQFHPGWN 658  
 DB 1886 -----NAPIRDPAAAPPVVTPTGPTLPLPNNSLFTLHPDAATLITDPRTIGRP 1935  
 QY 659 LTDALVIDNQLGHTNPPEHILDMHGLSLALQEGKLYHFDQLT--KGWGTAECDKO 716  
 DB 1936 YTSADTQLHALGDHT---LHRLGDGYEQLIRE-----QLAQLTGRRLRDLGYTDDQO 1988  
 QY 717 LKKGLDGAAYLLKDGVEK-----RLNINQSTSSI-----KHGTENFVSLPHVRN 760  
 DB 1989 YRALLDAGVTAVAKHQLRGIALSADQLAQLTSDIVWLQVQDQLPDGTTTRALVPRLYL 2048  
 QY 761 KPEPGDALOGLKNDKQAQAMAVIGNKYLALTEKGDIRSFQIKPGTQQLERPAQTLRSRG 820  
 DB 2049 RPTGTD---LTPDGALLAAASTTINAH-TFTNTGTIDARHL-----IDINAHMTDQOQ 2097

QY 821 -----ISGELKDIHVDHKQNLIALYTH--EGEVFHP--REAWQNGAESSWHL- 865  
 DB 2098 GRLTADAIIHTTGDPTTLGGOFKARGYLKVHAGQNFSLASTLRDATTOGTRHHSVTELD 2157  
 QY 866 -----ALPOSESKLSLSDMSHEHKPIATFEDGSOHOLKAGGWHAYAAEPERGFLAVGT- 917  
 DB 2158 QOAGFTVTCPGAYLGL-STDQAMTHNGVAINNTCTD-----GYTSLNA--TGPLHLGLT 2208  
 QY 918 -----SGSQTVFNRLMOCVKGKVIPIGSGLTIVKLSA----- 947  
 DB 2209 TTHRSDDTTQWDPRNSRHSRIDTEYGTSTIG-NGDIQLNSQDINLRAATLHSTOGTITAL 2267  
 QY 948 QTGGMT-----CAEGRKVSSKSERIRAVAFNFTMTSPRIKNAAY 988  
 DB 2268 ATGNVTIITHGDTIOYTSODSKTRKSLNLSRTTTTTHADQOQTAIGSTLSADKVFVKGN 2327  
 QY 989 ATQHWQOGRGLKPLIEMOGALIKOLDAHNVR-----HNAPODQLQSKLETLDLGEHCAEL 1044  
 DB 2328 ITVTGSHVVSADAGTYMQAEHDLTLQAATHTTQSTYSHHTKQRL-----IRNGGASL 2379  
 QY 1045 -LNDMKRFRDELEQSATRSVTVLGQHG-----VLKSNGEINSEFPKSP 1087  
 DB 2380 TLGNOSQRTDSTTTATTTTGLIGATNGNVNLLAGCHYQOIGSDVLSPHGDIDHAKKVD 2439  
 QY 1088 GKALVQSENVNRSQDL-----SKSLOQAVHATPPSAESKLOSMLGH 1129  
 DB 2440 ---ITQAHHTSOTTOHTATROSLTVALSTPLIAGATAOQMOHAAARSQDRLQALAGL 2496  
 QY 1130 FVSACVDMSHQGEIPLGRQRPNDKTALTLSRLILDTVTIGELHELADKAKLVSDHKPD 1189  
 DB 2497 TIALGA-----KNTIDAVRQ--DPRALGGLNAS-----LTVG----- 2526  
 QY 1190 ADQIKOLKQOFTLREKRYESNPVHYTDMGFTNKALEANYDAVKAFINAFKKEHGVN 1249  
 DB 2527 -----RSTHDSSTTTTSTAAGSNVTAGGNVH---ISATGDCGTASTLTIGSDVRG-- 2574  
 QY 1250 LTRTVLESQGSAAELAKLNTLLSLDSGESMSFSRYG-----GGVSTVFVFTLS 1300  
 DB 2575 -DTMTYLKADGDIAL--AAQNT-----TWQRNRRGSAGVGVAVNLGSGGTSAGLTAHAS 2628  
 QY 1301 KKVPVPVPIPGAGITLDRAYNLSFRTSGG--LNVSPGRD-----GGVSGNI 1344  
 DB 2629 TST-----GSGHSTDLTW--SNSHVGGCNLLAIDAGDGLLMKGAICTAKHVIADIAGNL 2680  
 QY 1345 MVATGHDVMPYMGKTSAGNASDWLSAKHKISPDLPRIGAANSVGTLOGLTIONSCLKFLE 1404  
 DB 2681 RIESQDTHQVRSRDSRLGGS-----LTAGAGFSGSANLHNHOTIRSDYASV 2726  
 QY 1405 DELPGF-----IHGLTH--GTLTP-AELLQKTEHOMKOGSKLTFSDTSAN-- 1448  
 DB 2727 TEQSGLFTGDDGYQLTVGGTHLIGGAITSNSTAHNGL-NSLDTGLILIONENHANYT 2785  
 QY 1449 ---LDLRAGINLNE-----DGSKPNGVYVARVS----- 1472  
 DB 2786 ATQVNLGGYSRNGCTVGTDOQGHAAATATQVPGTTLPSHNGLSASPPSAMTARDSSHST 2845  
 QY 1473 -AGLSASANLAAGSERSTTSQGFSTTSASNRP----- 1506  
 DB 2846 YSGISQSGALTIRDDTAQHALTGHTAAETIATLNRDLTDTATSNALTPIFDEQIRINAGFD 2905  
 QY 1507 -----TFLNGVAGANL-----TAALGVASHSTHEGKPVGIFPAFTSNVSAALA 1551  
 DB 2906 IVSGLQRTGTFFINRAAEADLUKTRQATAAADHAADPSN-----GFNDQOQRTLRDOAIA 2960  
 QY 1552 LDNR-----TSQSISLELKRAPVTSNDISELSTILGKHF-----KDSATT 1592  
 DB 2961 LTNEAHLKDDAWGPGTYRQITTAAG---ASGNVSAASSDLAKHMIYNYVOOQATAI 3017  
 QY 1593 KMLAALKELDDAKPAE-QJHIL-----QOHFSA-----KDV 1622  
 DB 3018 CHWVATGOLTEGSPHLAALHALLACAGAAASQHCSSGAQAAASSVLTLGLFSDPRPEDI 3077

QY 1623 VGDYERAVRNKKLIVIRQQAADSHSMELGSGASHSTTYNNLSRINDGI-----VELLHK 1677  
 Db 3078 AQDR--EAKRNLTISVIGIASIGHT-DEPATATHA-----IAAVDNWLAQYVQMLNE 3130  
 QY 1678 HFDAALPASSAKRIGEMNNDPALKDIIKOLOSTPSSASVSMELDKGLREQTEKAILDG 1737  
 Db 3131 EFEEA---TEKEGR-----EERKVRKAWRE-IDARQDKLTVDG 3166  
 QY 1738 KV-GREEGVLFODRNNLR--VKSVSVSOSVSKSGEFTNPALLLGTSTNSAAMSERNIG 1793  
 Db 3167 LKGLKESGI--NDINGLEHFLHPDVTVHELGI--LTHPTLLQLGESAVALLENKVS 3222  
 QY 1794 TIN--FKYQDQNTPR-----RFTL-EGGIAQANPOVASALTDLKKREGLE 1835  
 Db 3223 RMSEALLVGGDOHAQOLGEDLGSLVADVGFALAAAGTAKAAEILGEAGISLSKDVLE 3279  
 RESULT 7  
 Q9PBE8 PRELIMINARY; PRT: 3442 AA.  
 AC Q9PBE8:  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE HEMAGGLUTININ-LIKE SECRETED PROTEIN.  
 GN XF2196.  
 OS Xylella fastidiosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;  
 OC Xylella.  
 OC NCBI\_TaxID-2371;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-9A5C;  
 RX MEDLINE-20365717; PubMed-10910347;  
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,  
 RA Alvaranga R., Alves L.M.C., Araya J.E., Bala G.S., Baptista C.S.,  
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,  
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,  
 RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,  
 RA Coutinho L.P., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,  
 RA Facinani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,  
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Fulan L.R.,  
 RA Garnier M., Goldman M.H.S., Goldman M.H.S., Gomes S.L., Gruber A.,  
 RA Ho P.L., Hohelsel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,  
 RA Krieger J.E., Kuramae E.E., Lagret F., Lambais M.R., Leite L.C.C.,  
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,  
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,  
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.F.,  
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,  
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,  
 RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,  
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,  
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,  
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,  
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,  
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,  
 RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,  
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,  
 RA Vallada H., Van Sluys M.A., Verjovsky-Almeida S., Vettore A.L.,  
 RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;  
 FT "The genome sequence of the plant pathogen Xylella fastidiosa."  
 RL Nature 406:151-157(2000).  
 DR EMBL: AE004032; AAF84995.1; -  
 DR INTERPRO: IPR000267; -  
 DR INTERPRO: IPR001424; -  
 DR PROSITE: PS00087; SOD\_CU\_ZN\_1; UNKNOWN1.  
 DR PROSITE: PS00144; ASN\_GUN\_ASE\_1; UNKNOWN1.  
 SQ SEQUENCE 3442 AA; 360148 MW; AAE30CDE923E3C6E CRC64;

Query Match 2.6%; Score 247; DB 2; Length 3442;  
 Best Local Similarity 18.5%; Pred. No. 4.4e-05;  
 Matches 412; Conservative 292; Mismatches 815; Indels 708; Gaps 103;

QY 3 LKSLGTGHEAAVHTAAHNVPVGHVALQQGSSSSPQAAAALAAEGKNRGMRIHQST 62  
 Db 1374 IDNLGTGRGLYGDHIAL-----HAQTLNRDETSDGHTHAATIAARQLDIDGADTLRNAN 1428  
 QY 63 A---ADGISAHQKKSFSLRGCL---GTKKF---SRSAQOGOPGTHHSGKATRLRLDARD 114  
 Db 1429 AMILSDGDAI-----GATLDNALHATGATATLLDNRSATIDITGTLNITTTTNNI--RD 1481  
 QY 115 DGETOHEAAPDA---ARLTRSGVKKRRNMDMAGRPVWKGSGGDEKVPYQKRRHOLNMF 171  
 Db 1482 N---VHIAHAPDVVTEARMEQPHW---RKNQPN-----GGSGNFRTSNVDAHDI--- 1525  
 QY 172 GOMQRTMLSKMARPASANAGDRLQHSPPHIPGSHHEIKEEPPVGSTK-----A 219  
 Db 1526 -----YYLNPADI IKDD-----PYITPDQQIHRATVRLTPQTSVAFYARGGLYA 1570  
 QY 220 TTAHADRVETIAQEDDDSEFQQLHQOQLARENPQPQKLGAVATPISARFQPKLTAVAES 279  
 Db 1571 SOAERRRMDLTARTGDSVVLYYTD-----QDKQPNPDHVA---ATNHS 1613  
 QY 280 VLEGDTTQSPKPLKQSMKLGSGAGVTPPLAVTLDRKGLQAPDPNPPALNTLLKOTLGKDTQ 339  
 Db 1634 AFIGLTPQONERFQT-----VPITYAPGDDRLTYDSNG-----TCTDDCV 1655  
 QY 340 HYLAAHSSDGSQHLDDNKGHLFDIKSTATSYVLNHNHPGEIKGLAQAGTGSVSDG 399  
 Db 1656 RLVTWHDYTD-PDHTLID-----MHRG-PNDVR-----DN 1683  
 QY 400 KSGKISLGSCTQSHNKTMLSQPGEAHRSLLTGIIQHPAGAAPGEGESIRLH-----D 451  
 Db 1684 EKYR-----DATRTQODILNDPAGAPALIQTG-----GAMMIQTDLRNHHVADLLAGD 1733  
 QY 452 DKIHILHPGLVWQOSADKTHSQLSRQADGKLYALKDNRTLQNLSDNKSSEKLVDKIKSY 511  
 Db 1734 QIVGLPPTKEKSDEHKY-----KRVLLIDNRAQ-L-LS-----RTDTFNI 1776  
 QY 512 SVDQRQVA-----ILDTDPGRHKMSIMPSLDASPESHISLSLHFADAHGOLLHG 561  
 Db 1777 STTYRGKVPSEWSESRTTPTTQIGGRITSGGHIAAQAQLNNVNTDSTHAPEIQHLYTN 1836  
 QY 562 KSELEAQSVAISHGLRVVADSEGLFSAAPKPGQGNELKMKAMPQHALDEHFGHDHQLS 621  
 Db 1837 PS---TQTLSVWNGVITVTDTSPSLHTVSLADNPGSAGQELTYIPDQSLTTP----- 1885  
 QY 622 GFHDDHGOINLVK-----NN-----FRQOACPLGNDHQHFGW 658  
 Db 1886 -----NAPIRPAAPAVTPTGPTLPLPNSLFTLHPDAATLITTPDRTLGRP 1935  
 QY 659 LTDALVIDNQLGHHNPEPEILDMGHLSLALQEGKLHYFDQLT---KGWTGAESDCQ 716  
 Db 1936 YTSADTQLHALGDHT---LHKRLGDCGYEQRILIRE-----QLAQLTGRRLDGYTDDQ 1988  
 QY 717 LKGLDCAAYLLKDEGVK-----RLNQSTSSI-----KHGTENVFSLPHVRN 760  
 Db 1989 YRALLDGAVTVAKOHLRPGIALSADQLAQTSDIVLWVQODVQDPDGTTRALVPRYL 2048  
 QY 761 KPEPGDALOGLNKDKRAQAMAVIGVKNYKALTEKGDIRSFQKPGQOOLERPAQTLRSRG 820  
 Db 2049 RPTGD---LTPDGALLAAASTINAH-TFTNGTIDARHL-----IDINAHTMDQOG 2097  
 QY 821 -----ISGELKDIHVHKQNYALYTH-BGEVPHQP---REAWQNGAESSSWHKL- 865  
 Db 2098 GRLTADAIHIHTTGDEFTLGGQKARGYLKVAQGNFLASLTDRATQTGRHHSVTELD 2157  
 QY 866 -----ALPQSESKLSLDMSEHKPLATPDECSQHLKAGGWHAYAAPERGSLAVGT- 917  
 Db 2158 QOAGFTVTGPGAYLGL-STQAMTHNGVA INNTGTD-----GYTSLNA--TGPLHLGTL 2208  
 QY 918 -----SGSQTVFNRLMQGVKGVPGSGLTVKLSA----- 947  
 Db 2209 TTHRSDDTQWDPRNSRHSRIDTEYGTISITG-NGDIQLNSGQDINLRAATLHSTQGTITAL 2267



```
QY 948 QTGGMT-----GAEGKVKSKFSERIRAYAFNPTMTSPRIKNAAY 988
Db 2268 ATGNVTITHGDTIOYTSQDSHTKRSGLLSNRSTTTTHADQOOTOAIGSTLSADRVFYKGN 2327
QY 989 ATOHQWQREGKLPYEMOGALIKOLDHNVR-----HNAPQPDLOSKLETLDLGEGHAEL 1044
Db 2328 ITVTGSHVVSADAGTYMQAEHDLTLQAAATHTTQSTYSHTTKORGL-----IRNGCASL 2379
QY 1045 -LNDMKRPFDELEQSATRSVTYVLQHQG-----VLKSNCEINSEFKPSP 1087
Db 2380 TIGNQSORTSTTTATTTTGSGLIGATNGVNTLLAGHYOQIGSDVLSPHGDIDIIHAKVD 2439
QY 1088 GKALVQSFNVRNSGDL-----SKSLQAAVHATPPSAESKLQSLMGH 1129
Db 2440 ---IQAHTTSQTTQHTATRSGLTVALSTPLIAGTAQTAQOQHAARSQDPLQALAGL 2496
QY 1130 FVSQVDMHQGEIPLGRQRPNDKTAITSRLIITVTIGBELHELADKAKLVSDHKPD 1189
Db 2497 TTALGA-----KNTIDAVRQ-DPRALGGLNAS-----LTVG----- 2526
QY 1190 ADQIKOLRQOFTLREKRYESPVKHYTDMGFTTHNKALEANYDAVKAFINAFKKEHIGVN 1249
Db 2527 -----RSTHSTTTTTTSTAAGSNWTAGNVH-----ISATGDTASTLTIOGSDVRG-- 2574
QY 1250 LTRTVLESQGAELAKKLNTLLSDSGESMSFSRSYG-----GGVSTVFVPTLS 1300
Db 2575 -DTMYLKAQGDIAL--AAQNTV-----TWQRNRRGSAVGVAVNLGSGGTSAGLTAHAS 2628
QY 1301 KVPVPVPGAGITLDRAYNLFSRTSGG--LNVSPGRD-----GGVSGNI 1344
Db 2629 TST-----GSGHSTDLTW--SNSHVGGGMLAIDAGGDLMLKGAIGTAKHVIADIAGNL 2680
QY 1345 MYATCHDVMYPTGKTSAGNASDWLSAKHISPDLRIGAAVSGTLOGTILONSLKPKLTE 1404
Db 2681 RIESLODTHQYRSTDRKSLGS-----LTAGAFSGSANLNLHQTIRSDYASV 2726
QY 1405 DELPGF-----IHGLTH-----GTLTP-AELLQKQIEHQMKQSKLTFVSVDTSAN-- 1448
Db 2727 TQSGLGFTGCGYQVLTGCGOHLGCAITSNSTAIHNGL-NSLDTGTLLIQLNIENHANYT 2785
QY 1449 ---LDLRAGINELNE-----DQSKPNQGVTVARVS----- 1472
Db 2786 ATQVNLGGYSRNGTGTGTDQOHAATATQVPGTTLPSHNLGSLASPPSAMTARDSSHSTT 2845
QY 1473 -AGLSASANLAAGRSRSTTSOGFTSTASNNRP----- 1506
Db 2846 YSGISOGALTIRDDTAQHALLGHTAAETIATLNRLLDTATSNALTPIFDEORINAGFD 2905
QY 1507 -----TFLNGYGAGANL-----TAALGVAH-----SSTHEG 1532
Db 2906 IVSGLORETGTFINNRAEADLKTQATAADHAAHDPSNCFNQORHALBEQAIALTNEA 2965
QY 1533 KPV--GIFPAFTSTNVSAALADNRQTSQISLSELEKRAEPTVSNDSIELTSTLQKHF---- 1586
Db 2966 HAIKDAWPGGYRQITTTALAAG--ASGNVS-----AASNVSAAASDLAKHMTVNY 3015
QY 1587 ---KDSATTKMLAALKELDDAKPAE-QLHIL-----QQHFA----- 1619
Db 3016 VOQOGATAIGHVWATVQLTEGSLPHAAHALLACAGAAASQOHCSSAQOAGAAASSVLTGL 3075
QY 1620 -----KDVVGDERVEAVRNKLKLVIRQOQAADSHSMELGSAHSTTYNNLSRINNDGI-- 1671
Db 3076 FSDPRPDTQDR--EAKNKLITSIVTGIASTGNT--DAATATHAA-----IAAVDNWNLAA 3128
QY 1672 ---VELLKHKFDAAIPASSAKRIGEMMNNDPALKDIIOQLQSTPPFSSASVMELKQGLRE 1728
Db 3129 KOYVQMLNEFEAA---TEKEKGRL-----EEEKVRKAWRE-IDA 3164
QY 1729 QTEKAILDQKV--GREEVGLFQDRNLR--VKSVSYSOSVSKSEGFENPTALLGTSNSA 1784
Db 3165 RODKLTVDGLLGLKRESGI--NDINGLEHLFLHPVDVTVHELGI--LTHPTLLQLGESA 3220
QY 1785 AMSMERNIGTIN--FKYGDQNTPRFTLEGG--IAQANPQVASALT-----D 1828
```

```
Db 3221 VOELLNKVRSEALLVGGDQHA-QQFGEDLGSVIADVGVALAAGTFFKAAEILGAGIN 3279
QY 1829 LKKEGLE 1835
Db 3280 LSKDVLE 3286

RESULT 8
Q9K0T0 PRELIMINARY; PRT: 2703 AA.
AC Q9K0T0;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE HEMAGGLUTININ/HEMOLYSIN-RELATED PROTEIN.
GN NMB0493.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGROUP B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., Deboy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Ciftone H., Clark E.B.,
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Masignani V., Pizza M., Grand G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58.";
RL Science 287:1809-1815(2000).
DR EMBL; AE002405; AAF40927.1; -.
DR TIGR; NMB0493; -.
SQ SEQUENCE 2703 AA; 288263 MW; 214537DIF261F00A CRC64;
```

```
Query Match 2.6%; Score 245; DB 2; Length 2703;
Best Local Similarity 19.0%; Pred. No. 3.9e-05;
Matches 414; Conservative 282; Mismatches 766; Indels 712; Gaps 105;

QY 10 HKAAVHTAAHPVHGVALQOQSSSSSPQAAAAAASAAEKNCKMKEPRIHPSTAAD--G 66
Db 573 HDLAVNTQAKNSGH-LLTQTKIDNRELHNAGETAA-----NNLTIHSGRLSNDKKN 626
QY 67 ISAAHQKKFSLR-----GCLGKKFSRQPOGPGTTHSKGATLRDLARDGGE 117
Db 627 IRAAHLQDLDXAGLHNAGNLADSGVTYTKNNLRNTGKVSARLNTGOTLIDNTRGRIAE 686
QY 118 TQHEAAAAPDAARLTRSGG-----VKRRNDDMDAGRPM-----VKGG---SGE 156
Db 687 TVN-----IQSQQLTNQSGHITATEQITINSRVNDQNGKLLSANAQAVSDGLYNQHG 742
QY 157 DKVPTQQRHQLNFPQMRQTHMLSKMAHPASANAGDRLOHSPPHIPGSHHEIKPEPVST 216
Db 743 IATNRQLSIHDKN-----QNTLALNNADGTIQSAGN-----VSLQAKSLA 782
QY 217 SKATTAAHADRVEIAQEDDDSEFQOLHOORLARENPPOPKLGVATPISARFQPKLTAV 276
Db 783 NNGLTAGNKLDIAITDD-----FVVERD-----LTAGKQLNLS-- 816
QY 277 AESVLEGTDTTQSPKLPQSMKLGSGAGVTPLAVTLDDKGLQLAPADNPPLNTLLKQTLGK 336
Db 817 IKGRLNKNTHTLQ-----AGHT---LKLNAGNI---DN-----QVTGK 847
QY 337 DTQHYLAHASDSQSQHLLLNKNKHLFDIKSTATSYSLHNSHPGEEKKLAQAGTGSVS 396
Db 848 -----IIGGEQTDITSEQHVNRG-----LINS-----DGLTHIGAGQTL 882
QY 397 VDGKSGKLSLGSQTSQSHNKTMLSQPGEAHRSLITGTWQHPAGAAARPOGE--SIRLHDDKI 454
```

Db 883 TMTGKI-YGNHIALDAQILLNRETTETGSTRKAG-----AIAARKLDIGAKEIHNQEG 936  
QY 455 HILPE--LGWWSADKTHSO-----LSROADGKLYALKDNRTLQNLSDNKS 500  
Db 937 ALLSEGI FAVGNRLDEQHAAGMADTFVNGSAGLEVQGD---ALMSVRMOMNHNHF 992  
QY 501 SE---KLVDKTKSYV-----DORGOVAILDTFGRHKMSIMPSLDASPESHIS-L 549  
Db 993 TETYLAKAEQVRDVTVLGONTYYQAGDKGLFDNSQOQ-KDQTTATPHLKNGSRLEANOW 1051  
QY 550 HPADAH-----OGLLH-----GKSELEAQSAVSHGRVLVAD----- 581  
Db 1052 HVYDHIETIKERIENRPAHITVGGDLTASGQWLNKDSRIYVGGRIITDDLKQKEITN 1111  
QY 582 --SEGLFSAALPKQGD-----GNEKMKAMPOHALDEHFGHDHQLISGRPHDDHQ 630  
Db 1112 QSTTGKRTDAVCTQWDSVTKKGWYSGRKRQRTERNHTP-----YHDTQL--FTHDFTDP 1165  
QY 631 LNALYKN-----NFRQOHAC-----PLGNDHQFHPG--WNLTDAVLIDNQLGLHHTN 675  
Db 1166 VSVIQNAASPSPQPAASAIAKLIDGVSTAAVNGQRITHTGNVSLNNAATVTLPNSSLVTH 1225  
QY 676 PEPHEIDMGLSLAQEGKLYHFDQTKGTGAESDCKQLKKGLDCAAYLLKDGVEKR 735  
Db 1226 PD-----NKGW-----LVETDPQFADY--RRWLGSYMLQQLQDITNHLHKLRLGDDGYEQ 1273  
QY 736 LNIQNTSSIKHGTENVFSLPHVRNKPPEPGDALQGLKDKQAKAMAVIGNKYLALTEKG 795  
Db 1274 KLVN-----EQIHQLTGYRR-----LDGYSDEE-----QFKALMDNG 1306  
QY 796 --DIRFOIKPG---TOO-----LERPAOTLSREGISGELKDIIHVDHKONLYAL 839  
Db 1307 LTAATFTGLTPGIALSAEQVARTLSDIVWMENQTVLTS-----DGSTQTVLV---PKVVAL 1359  
QY 840 THEGEV-----PHQPREAWNGAESSE-----WHLKALPQSES----- 872  
Db 1360 ARKGDLNTSGGLISAEQVLLKQLONGLNTSGTITAGQAVLIQIARNINSNGNIQADQIGLK 1419  
QY 873 KLKSDMSHEHKPIATFEDGSOHQ-----DGQVQAGRLTLTAQAQNLNLTQTSNGERNGNTAIDRMAG 1468  
Db 1420 AEKSIINI-----DGGQVQAGRLTLTAQAQNLNLTQTSNGERNGNTAIDRMAG 1468  
QY 912 PLAVTSSGQTVFNLQMKGKVIIPSGSLTVKLSAQGTGWTGAEGKVSSEKSE---RI 968  
Db 1469 INVVG-SHTEQVDNRTSDGI-----LSLHASNDINLNAATVSNQVKDGTQI 1514  
QY 969 RAYAFNPTMTSPRIKNAAYATQHGQWQREGKPLVEMOGALIKQLDANVRHNAPODL 1028  
Db 1515 TA-GNNLNLGTIRTHERAYGT-----LDDENHRHVRSQSTEV 1550  
QY 1029 QSKLETLDLGEHGAELL--NDMKRFRDELEQ-----SATRSVT-----VL 1066  
Db 1551 GSSIRT---QNGALLRAGNDLKIROGELEAEGBGKTVLAAGRDVITSEGRQITELDTSVS 1606  
QY 1067 GOHQGVLSKNGEINSEFKPS-----PGKALV---QSENVN-----RSG 1101  
Db 1607 GKSKGILSST--KTHDRIYRSHDEAVGNSGTGGKMIYVAAGODIINVRGNSLSDKGIVILKAG 1665  
QY 1102 QDLKSLOQAOVHATPPSAESKILQSMGLHFVSAGVDMSHQKEIPLGRQR--DPNDKLTALT 1159  
Db 1666 HDIDISTAHNRYTGNEYHESKSGVNG---TGGLGFT-----IGNRKITDDTDRINIV 1715  
QY 1160 KSRLLJ-----DVTITIGELHELADKAKLVSDHDKPDADQIKOLQOQFTTLREKRYESNPVK 1214  
Db 1716 HTSGTIGSLNGDVTVAGNRYRQTGSTVSSPEGRNTVTAKSIDVEF---ANNRYATDYAH 1772  
QY 1215 HYDMGFT---HNKALEANYDAVKAIFNAFKKEHHGVNLTTRIVLESQGSAAELAKLKNT 1271  
Db 1773 TOEQKGLTVALNVPVQAAQNEIQAAQNVGKSKNRVNAANAANAQW--SYQATQOQMOQF 1831  
QY 1272 LLSLDSGSMSPSRSYGGGVSTVFPVTLTK-----KVPVPVPGAGITLDRAYNLFS 1323

Db 1832 AFSSSAGQONNOSPISVIITYGEOKSRNEQKRHYTAAASQIIGKQT-----TL 1884  
QY 1324 SRTSGG---LNVSFGRDGGVSGNIMVATGHDVM--PYMTGKKTSGAGNASDW--LSAKHKI 1376  
Db 1885 AATGSEQSNINITGSDVIGHAGTALIADNHIRLOSAKQDQSGEQSKNSGNGAVAVKI 1944  
QY 1377 SPDLRTGAASVGTL-QCTLQNSLAKFKLTEDLPFIHGLTH-----GTULT 1420  
Db 1945 GNGIRFGITAGNIGKGEQGG-----STTHRRHTVGTSTGKTIIRSGGDTTLK 1993  
QY 1421 PAELLQKGE-----HMQKQ-----GSKLT--- 1440  
Db 1994 GVOLICKGQADTRNLHIESVQDQTTETYSQKQONGNVQVTVGYGFASGYRSYKSKADHA 2053  
QY 1441 -----FSVDTLSANLDRAGINLNDGSKPENGVTARVASGLSASANLAAGSR 1486  
Db 2054 SVTQSGSIYAGEDGYQIKVRDNTDLKGGIITSSQSAEDKGNLFQTAITLTA--DIONHSH 2112  
QY 1487 ERSTTSQPGS-----TTSASNNRPTFLNGVGAG-----ANLTAALGV---AH 1526  
Db 2113 YEGRSFSGIGSFDLNGWDGTVDKQGRPTDRISPAAGYSGDGSKNSTTRSGVNTNTH 2172  
QY 1527 SSTHECKPVCIFPAFTSTNVSAALADNRTSQSISLELKA--EPVTSNDIS---ELTST 1581  
Db 2173 IDEAGQLARTGTAKETEARIYTGIDTADQSHGHLKNSFDKDAVAKELNLOREVTK 2232  
QY 1582 LQKHFKDSATTMLAALKELDDAKPABOLHILOQHFSAKDVVDERYEAVRNKLKLVIRQ 1641  
Db 2233 FGRN-----AAQAAV-----ADKLGTQSY-----ERYQEARLT--LEAEL 2268  
QY 1642 QAADSHME-----LGSASHSTTNNLSRINN 1668  
Db 2269 QNTDSEAKAAFRASLGQVNAVLAENQSYDWTKEGGIGRSILHGAAGLTTGSLGILA 2328  
QY 1669 DGIVELLHFKHFDALPASSAKRGLGEMNDPALKDIIKQLOSTPFSSASVSMELKGLRE 1728  
Db 2329 GGTSLAAPYLDKA-----AENLG-----PAGAAVNALGAAIYAT----- 2366  
QY 1729 QTEKATLDKGVGEEVGLVFQDRNN--LRVKSVSVS-----QSVSKSEGFNTPALLL 1778  
Db 2367 -----GSGGCAVVGANV--DWNRRQLHPKEMALADKYAEALKREVEKREGRKI----- 2412  
QY 1779 GTSNSAAMSERNI 1792  
Db 2413 -SSQEAAMRIRQI 2425

RESULT 9  
Q14789 PRELIMINARY; PRT; 3259 AA.  
ID Q14789; Q14398;  
AC Q14789; Q14398;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
DE GIANTIN (GCP372) (MACROGOLGIN) (GOLGI AUTOANTIGEN, GOLGIN SUBFAMILY B, 1)  
DE GOLGB1.  
GN Homo sapiens (Human).  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=009606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94187728; PubMed=7511208;  
RA Seelig H.P., Schranz P., Schroeter H., Wiemann C., Griffiths G., Renz M.;  
RT "Molecular genetic analyses of a 376-kilodalton Golgi complex membrane protein (giantin) [retracted in Mol Cell Biol 1995 Jan;15(1):591].";  
RL Mol. Cell. Biol. 14:2564-2576(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94257116; PubMed=8198703;  
RA Seelig H.P., Schranz P., Schroeter H., Wiemann C., Griffiths G.,



RA Renz M.;  
RT "Macrogolgin--a new 376 kD Golgi complex outer membrane protein as  
RT target of antibodies in patients with rheumatic diseases and HIV  
RT infections."; J. Autoimmun. 7:67-91(1994).  
RL [3]  
RN  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95100974; PubMed=7802676;  
RA Sohma M., Misumi Y., Fujiwara T., Nishioka M., Ikehara Y.;  
RT "Molecular cloning and sequence analysis of a human 372-kDa protein  
RT localized in the Golgi complex."; Biochem. Biophys. Res. Commun. 205:1399-1408(1994).  
RL Biochem. Biophys. Res. Commun. 205:1399-1408(1994).  
CC -!- FUNCTION: MAY PARTICIPATE IN FORMING INTERCISTERNAL CROSS-BRIDGES  
CC OF THE GOLGI COMPLEX.  
CC  
CC -!- SUBUNIT: DISULFIDE-LINKED HOMODIMER.  
CC  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. GOLGI MEMBRANE.  
CC  
CC -!- ALTERNATIVE PRODUCTS: DIFFERENT ISOFORMS MAY ARISE BY ALTERNATIVE  
CC SPLICING.  
CC  
CC -!- DISEASE: ANTIGEN IN CHRONIC RHEUMATOID ARTHRITIS AND IN THE  
CC AUTOIMMUNE DISEASE SJOEGREN'S SYNDROME.  
DR EMBL; X75304; CAA53052.1; -;  
DR EMBL; D25542; BAA05025.1; -;  
DR MIM; 602500; -;  
KW Golgi stack; Antigen; Coiled coil; Transmembrane;  
KW Alternative splicing.  
FT DOMAIN 1 3235 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 3236 3256 POTENTIAL.  
FT DOMAIN 3257 3259 LUMENAL.  
FT DOMAIN 48 110 COILED COIL (POTENTIAL).  
FT DOMAIN 127 223 COILED COIL (POTENTIAL).  
FT DOMAIN 238 448 COILED COIL (POTENTIAL).  
FT DOMAIN 460 526 COILED COIL (POTENTIAL).  
FT DOMAIN 545 593 COILED COIL (POTENTIAL).  
FT DOMAIN 677 956 COILED COIL (POTENTIAL).  
FT DOMAIN 969 1028 COILED COIL (POTENTIAL).  
FT DOMAIN 1062 1128 COILED COIL (POTENTIAL).  
FT DOMAIN 1154 1245 COILED COIL (POTENTIAL).  
FT DOMAIN 1301 1779 COILED COIL (POTENTIAL).  
FT DOMAIN 1828 2781 COILED COIL (POTENTIAL).  
FT DOMAIN 2797 2857 COILED COIL (POTENTIAL).  
FT DOMAIN 2872 2993 COILED COIL (POTENTIAL).  
FT DOMAIN 3026 3102 COILED COIL (POTENTIAL).  
FT DOMAIN 3133 3185 COILED COIL (POTENTIAL).  
FT DOMAIN 2420 2423 POLY-GLU.  
FT DOMAIN 2993 2996 POLY-SER.  
FT DOMAIN 1 39 MISSING (IN REF. 3).  
FT VARSPLIC 215 215 A -> AQLSSM (IN REF. 3).  
FT VARSPLIC 1765 1765 D -> G (IN REF. 3).  
FT CONFLICT 2950 2950 H -> D (IN REF. 3).  
SQ SEQUENCE 3259 AA; 376075 MW; 60376A20D8A178DD CRC64;

Query Match 2.6%; Score 244.5; DB 4; Length 3259;  
Best Local Similarity 18.6%; Pred. No. 5.7e-05;  
Matches 389; Conservative 311; Mismatches 816; Indels 577; Gaps 92;

Qy 23 GHGVALQGGSSSSPQNAASLAAGKNGKMPRIHQ--PSTAADGISAHQKKSFSRLR 80  
Db 603 GEGIA-----PIKKVFLDTGQDPPLMPNESSLPAVEKEQAETHOSRTSEEI- 652  
Qy 81 GCLGTKFSRPAQOPGPTTHSKGATLRDLLARDGDGTOHEAAPDAAR----- 129  
Db 653 -----SLNDAGVELKS--TKQGDG-KSLSAVDPDIGOCHQDELERLKS 691  
Qy 130 -----LTRSGGVKRRNMDMAGRMVKGSGEDKVPYTOQKRHOLN-----FGQMR 175  
Db 692 QILELELNFHKAQIEYKLNDEKA-----KEISNLNQLIEEPKKNADNNNSAFTALSEER 746  
Qy 176 QTMLSKM-----AHPASANAGDRLHSPPHIPGSHHEIKPEPVGSGTSK 218  
Db 747 DQLLSQVKELSMVTELRAQVKOLENLAPAEQRRLDYES-----QTAHDNLLTEQIHSLSI 803  
Qy 219 ATTAHADRVETIAQED--DDSEFOOLHQQLARERENPPQPPLKGV-----ATPISARFQP 271

Db 804 EAKSKDVKIEVLONELDDVOLQFSEQSTLIRLSOLOLNKSEVLEGAERVRRHISKVEE 863  
Qy 272 KLTAVAESVLEGTDTQSPKLP-----QSMKSGAGVTPLAVTLTDGKGLQALPDN-- 322  
Db 864 LSQALSQKELEITKMDOLLEKKRDVETLQOTIEEKDQOVTEISFMTKXVOLNEEKF 923  
Qy 323 -----PPALNTLLKOTGLKDTQHYLAHASSGDSQHLLDNKGHLFDIKSTATSYSV 374  
Db 924 LGVEIKTLKEQNLNLSRAEAKEKQVEDNEVSSGLKQN-----YDEMPAGQISK 974  
Qy 375 LHNHPGCI-----KGLAAQAGTGSVSDGKSKIS---LGSQTQSHNKTMLSQP-- 421  
Db 975 BELQHEFDLLKKENQRRKKLQAAALINRKELLQVRSEELANLKEELPLSETER 1034  
Qy 422 GEARSLLTGIWHPAGAARPOGESIRLHDDRIHILHPELGVWQSDAKDTHSOLSQAQD 481  
Db 1035 GEVEEDKENKEYSEKCVTSKQOEIYL---KQTISEKEVEL-OHTRKDLLEKLAEEQF 1090  
Qy 482 KLYALKDNRTLN-----LSDNKSSEKLVKDIKSYSD--ORGOVAILTDTPGR 528  
Db 1091 QALVKOMNTQDKTNQIDLQAEISENQA---IIQKLITSNTDASDGDVALVKET--- 1144  
Qy 529 HKMSIMPSLDASPESHISLSLHFADAHQGLLHGKSELEAOSVAISHGRLVWVADSECKLFS 588  
Db 1145 --VVISPCTGSE-----HWKPELEEKILALEKEK---EQLQKQLQE 1182  
Qy 589 A-----AIPKQGD-----GNELKMKAMPQHALDEHFGHDHQTSGFFHDDHGQNLALVK 637  
Db 1183 ALTSRKAILKKAQAEKERHLRELKQKDDYNRLQEQFDEQSKENENIGDLQRLQIQVRE 1242  
Qy 638 NF-----ROHACPLGNDHQFPGWNLTDALVIDNQLGLHHTNPEPHEILDMHGLSL 690  
Db 1243 SIDGKLFPSTQDQESC-----SSTFG--LEELPLFKATE--QHHTQP-----VLE- 1281  
Qy 691 ALQEGKLYFPQLTKGWTGAESDCKQLKKGLDGAAY--LLKQGEVKKR---LNNINOSTSS 744  
Db 1282 -----SNLCPDWPFSHSEDASALOGGTSVAQIKAKQLEAEKVELELVKVSSTSE 1331  
Qy 745 IKHGTENVFSLPHVKNKPEPGDALQGLNKKDKKAQAMAVIGVNKYLALTEKGDIRSFQIKP 804  
Db 1332 LTRKSEEVFOLEQEQINK-----QGL-----ETES--LKT 1358  
Qy 805 GTQOLERAPOTLREGISSELKDIIHVDHKNLYALTHEGEVPHQPRE---AWONGAESSS 861  
Db 1359 VSHEAEVHAESLQOKLESSQLOIAGLEHLRELQPKLDELQKLISKEEDVSYLSGOLSEK 1418  
Qy 862 WHKLA-----LPQSESKLSLSDMSHEKPIATFEDGSOHOLKAGWHAAYAAPERGLAVG 916  
Db 1419 EAALTKIQTEIEQEDDLIKALHTQLEMQ--AKEHDERIKQLQVCELMCKQKPEE---IG 1472  
Qy 917 TSGSOTVFNRLMQGVKGVIPGSGLTVKLSAQTGTMGAEGRKVSKFSERI---RAYAF 973  
Db 1473 EE-----SRAKQIQIRKL-----QAALISRKALKENKSLQEEELSLARGTIE 1514  
Qy 974 NPTMSTPRPIKNAAYATOHGQGRE-----GLKPLYEMOGALIKOLDAHNVHRHNAPOPD 1027  
Db 1515 RLTKSL-----ADVESQVSAQNKERKQTVLGRLLALQEEERDKLITMD-----KSLLENOS 1564  
Qy 1028 IQSKLETLDLGEHAELNDMKRFRDELEFQASATRSVTVLGHQGVLYKSNSEINSEFKPSP 1087  
Db 1565 LSSSCESLKALEG--LTEDKEKLVKEIESLSSKSLAESTEQ---EKIKELQKEY---- 1615  
Qy 1088 GKALVQSFNVRNSGDLKSLQAVAHATPPSAESKLSQMLGHFVSAGVDMSH-----QKG 1142  
Db 1616 -EILLQSY-----ENVSNEAERIOHVVEAVRQEK-QELYGLKLRSTEAANKKETEKQLQEA 1667  
Qy 1143 EIPLGRDRDPNDKTALTKSRLLIDTV-----TIGELHELADKAK-----LVSDHKPDQAO 1192  
Db 1668 EQEMEENKEMKMRKFAKSKQOKKILEEENDRLRAEVHPAGDTAKCECMETLSSNASKMKEE 1727  
Qy 1193 IKOLRQOQFDTLREK-----RYESNPVKHYVTDMGFTTHNKALEANYDAVAKFINAF 1241

```
Db 1728 LERVMEYETLSKKFQSLSEKSDSLSEEVQDLKHQIE-----DNVSKQANLEAT----- 1776
Qy 1242 KKEHGVNLTTRTVLESOGSALAKLKLNTLLSDS-----GESMSRSRSGGVSTVF-- 1295
Db 1777 EKHDQNTNVEGTOSIFGETE-----EODLSJMSMTRPCSESVPSAKSANPAVKDFSS 1831
Qy 1296 -----VPTLSKKVPVPVPIGAGITLDRAYNLSFRTSGG-----LNVSFGRDGGV 1340
Db 1832 HDEINNYLOQIDOLKERI-----AGLEEKQKNKEFSOTLENEKNTLLSQISTKDGEL 1884
Qy 1341 -----SGNIMVATGHVMPYMTGKTSAGNASOMLSAKHKISPDLRIGAAVGTGLOQ 1392
Db 1885 KMLQEVTKMNLNQQIOBELSRVTKLKETABEEKD-----DLEE 1924
Qy 1393 TLONSLKFKLFEDELPGFTLHGLTLPALLOKIEHOMKOGSKLTFPSVD----- 1444
Db 1925 RLNMOLA-----ELNGST-GNYCOVDTAQIKNEULESEMKNLKKVCSELEEKQQLVK 1977
Qy 1445 --TSANLDR-----AGININEDGSKPNGV-----TARVSAGLS 1476
Db 1978 EKTQVESEIRKEYLEKIQGAQ-KEPCNKSNAKELQELLKEQKQEVKQLQKDCIRYQEKIS 2036
Qy 1477 ASANLAAGSRERSTSGFGSTTSASNRPPTFLNGVGAGANLTAALGYAHSSTHEGKPVG 1536
Db 2037 A-----LERTVKALEFVQTESQKDLITKEN-----LAQAVEHRKKAQAEALS 2079
Qy 1537 IFPAFTSNVSAALAL-DN-----RTSQSISLELKRAEPTVTSNDIS-ELTSTLGHKF 1586
Db 2080 FKVLDDTQSEARVADNLNKLKKELOSNEKSVKSOMKQD-----EDLERLEQAEKHL 2135
Qy 1587 KDSATTMLAALKELDAPALHLOHFSKQVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGV 1646
Db 2136 KEKNK-----MOEKLDALRRKEVHL-----EETIGETQVTLNKKDEQVQQLQENLDS 2182
Qy 1647 HSMELG--SASHSTYNNLSRNNNGIVELLHKKHFDAAALPASSAKRLGEMNNDPALKDI 1704
Db 2183 TVTQAAFTKMSLQDDRDV-IDEAKWKFSDAISQSEKKEIRLKE--DNCVSLKQD 2239
Qy 1705 IKOLQSTPFSASVSM-ELKQGL-REOTEKAILDQKVGRE----- 1742
Db 2240 LRQM-----SIHMEELKINISRLHDKQIWESKAQTEVQLQKVCYDTLQGENKELLS 2291
Qy 1743 -----EVGVLFQDRNLRVSVSVSVSVSVSVSVSVSVSVSVSVSVSVSVSVSVSVSV 1788
Db 2292 QLEETRHLVHSSQNELAKLESLSKLDQTLNLSLEKCKEQK-----GNLEGIIRQO 2345
Qy 1789 ERNICTINFKYGO---DQNTPRRFT--LEGGIAQANPQVASALTDLKKEGLEM 1836
Db 2346 EADIONSFSEYQLETDLOASRELTSLRLHEEINMKKEQKIISLSLG-KEEAIOV 2397

RESULT 10
O76891 PRELIMINARY; PRT; 5327 AA.
ID O76891
AC O76891
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)
EG:49B4.1 PROTEIN.
GN EG:49B4.1.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA papagiannakis G., Siden-Kiamos I., Louis C.;
RT "Sequencing the distal X chromosome of Drosophila melanogaster.";
RN Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Benos P.;
```

```
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL031128; CAA20006.1; -.
DR FLYBASE; FBgn0025392; EG:49B4.1.
SQ SEQUENCE 5327 AA; 575942 MW; FEFEE23A118FF38A CRC64;

Query Match 2.5%; Score 240.5; DB 5; Length 5327;
Best Local Similarity 18.4%; Pred. No. 0.00022;
Matches 370; Conservative 271; Mismatches 803; Indels 567; Gaps 81;

Qy 18 AINPVGHVVALQOGSSSSPONAAASLAEGKNRKKPRIHOPSTAAAGI-----SAAH 71
Db 3517 ASRPTSVAESVDEAKESRESVAEKSLSLAK--EASRPASVAESVDEAKESKEE 3574
Qy 72 QOKKFSLGGCLGTTKFFRSAPQOG---PGTTHSKGATLRDLLARDDCGETQHEAAAPDA 128
Db 3575 SRRESVAEKSPLAKESRPAESVAESVAEKSKEYSRRESVAEKSPLPKESARPTSV 3634
Qy 129 -----RLTRSGGVKRRNMDM--AGRPVVGKGGEDKVPTQOKRHQLNFGOM 174
Db 3635 AESVKDEADKESRESGAESKPLASMEASRPTSVAESVKDETEKESRESRESVTEK 3694
Qy 175 ROTMLSKMAHPASA--NAGDRLQHSPPHIFGSHHIEKEEPVGSSTKATTAHADRVIAQE 232
Db 3695 SPLPSKEASRPTSVAESVKDEAK-----SKEESRESVAEKSPLAKESRSPASVAE 3747
Qy 233 DDDSEFQQLHQRLARERNPPOPKL---GVATPISARFOPKLTAVAESYLEGTDITQ 288
Db 3748 SIKDEAEGTKQE---SRRESMPESGKAESIKGQSSLASKETSRRPDSVSVESVKDETE--- 3801
Qy 289 SPLKPO-SMLKSGAGVTPPLAVTLDKGKLQAPNPALNTLLKTLQKDTQHYLAHHAS 347
Db 3802 ---KPEGSAIDKSOVASRSPESVAVS-----AKDEKSPHSRPSVADKSPD-----AS 3846
Qy 348 SDGSOHLLDNKGHLFDIKSTATS-----YSVLHNSHP-----GEIKKLAQAGTGSV 397
Db 3847 KEASRSL-----SVAETASPIEGEGRPSIADLSPLNLTGEAKGLP---TLSSPI 3894
Qy 398 DKGSKISIGSGTOSNKNTMLSGPGEAHRSLTGIOWHPAGAARPGQSGIRLHDDKIHL 457
Db 3895 DVAEGDFLEVKAESSRPAVLSKPAEFSQPDGTHTASTPVDPEASVLEIEVQO----- 3949
Qy 458 HPELVQWSADKDTHSOLSQAQDKLYALKDNR--TLQNLDNKSSEKLVDKIKSVYVDQ 515
Db 3950 HTSGV-----GATGATAETDLDLTETKSETVTVKQSETTLFELTTSKVES----- 3995
Qy 516 RGQVAILTDPGRHKMSIMPSLDASPEHSISLHFA-----DAHQGLLHGKSE--LEAQ 568
Db 3996 -----KVEVLESSVKQVEKQVTSVKQAEETTVDTSLEQLTKKSEQLTEIK 4041
Qy 569 SVAISHGRVLVADSEKGLFSAAI---PKQDGNELKMKAMPQHALDEHFGHDHQSIGFF 624
Db 4042 SVLDTN-----ISNVTNLFSTAVETIEKKVQDVTEKVKATEH-VSEHVTVTGESSTET 4095
Qy 625 HDDHGOLNALVKNNFRQOHACPLGN-----DHQFHPGWNLT 661
Db 4096 SQEKSLLDGLTFSELRETHITTVGSPETVTICERDEPVLDHIKEDEHRESPPSDVOK 4155
Qy 662 ALVIDNQLGLHHTNPEPEHILDMGHLSALQEQK-----LHYFDOLFKGWTGAESDCKO 716
Db 4156 AALIPPO-PMRPLSPREEVAKIVADVAKVLKSKDKIDIPDFDE-----RQ 4202
Qy 717 LKGLDGAAYLLKDGVEKVRNINOSTSIKHGTENVFSLPVHRNKNPCPDALQGLNKDDK 776
Db 4203 LEEKLSTADTEESDKSTRDEKSLEISVKVEISEKSPDQKSGPISTIEEDKIEQSEK 4262
Qy 777 AQAMAVIGVKNYLALTEKGDIRSFOIKPQTOOLERPACTLSREGISGELKDIHVDHKQNL 836
Db 4263 A-----QLRQGIILASSRP-----ESVASQESVSPSQS-- 4291
Qy 837 YALTHEGEVPHQPREAWQNGAESSSWHKLALPQS--ESKLKSLDMGHEHKPIATFEQSGH 895
Db 4292 -AASHE-----HKEVEL-----SESHKAEKSSRPESVASOVSEKDMKTSRPSAST---SQF 4338
```

QY 896 QKAGGWHYAAPERGLAVGTSGSOTVFNRLMOGVKGVIPGSLTVKLSAQTGGMTCA 955  
DB 4339 STKEGDEETES-----LLHSLTTTETVKOME-----EKSFESVTSVT 4380  
QY 956 EGRKVSXKFSERIRAYAFNPTMTPRPKNAAYATOHGWOGREGKLPKPLYEMOGALIKOLD 1015  
DB 4381 KSTVLSQSQTQVLRSESTSSLSKVEDSS-----RRSLSSLLAEKGGI----- 4427  
QY 1016 AHNVRHNAPOPDIOSKLETLDLGEGHGAELLNDMKRFRDELBOGATRSVTVLGQHQGVUKS 1075  
DB 4428 ATNTSLKEDTSASASOLE-----ELL-----VQSECSSESIVSIQTSIAOKS 4471  
QY 1076 NGEI-----NSEKFPSPCKALVQ-----SPNVNRSGOD- 1103  
DB 4472 NKEIKDARETKVTSQPTTTTSSATKDDSLKETVAEFLATEKIVSAKAEAFTEATKSADDC 4531  
QY 1104 LKSLQOAVHATPPSA-----ESKLQSLMGHFVSAGVDMHQKGEIPLGRQDRPNK 1155  
DB 4532 LKKTASAVSSTSASORALFVGTDESRRSLLSQASESRLTHSDPEEPAD---DVDER 4588  
QY 1156 TALTKSRILDTVTIGELHELADKAKLVSDHKPDADQIKOLRQOFDILREKRYESNPVKH 1215  
DB 4589 SSVKESR-----SKSIATI-----MMTSIYKPSDEM-----EPI-- 4617  
QY 1216 YTDWGFTHNKALPANYDAVKAFINAFKKEHGVNLTTRTVLES---QGSAEALAKKLKN--- 1270  
DB 4618 -----SKLVEEHEHVE---ELAQEVTSKTT-TLLOSSEQSTTSSTSKTGAS 4664  
QY 1271 -----TLLSLDSGESMFSRSYGGVSTVFPVTLKSKVPVPIPGAGITLDRAYNLSFSR 1325  
DB 4665 RVESITLTQMD-----QOTSOSQDPAKRTPT-----APVSPGV----- 4700  
QY 1326 TSGGLNVSGRGGVSGNIMVATGHVMPYTKTKTSAGNASDWLSAKHKITSPDLRIGAA 1385  
DB 4701 -----KAMSTG-----SAGSV-----ICAG 4716  
QY 1386 VSGTLOGTLQNSLKFTEDELPGFIHGLTHGLTTPAELLQKGIHQMKOGSKLTFVSVD 1445  
DB 4717 AGAVAAGGKCESSAAIVS-----SSGPMSP-----KDISGKSPGALTSQSOSI 4761  
QY 1446 SANLDRAGINLNEGDSKNGVTVARVSAGLSASANLAAGSRERSTTSQGFGSTTSASNRR 1505  
DB 4762 PTPLGRESHTDTPESPFPKPTSPRPYRSKDELKSLQEMOHHSQEQMAGAAAAAGACEGDI 4821  
QY 1506 PTF--LNGVGAGANLTAALGVAHSTHEGKPVGIFPAFTSTNVSAALALDNR---TSQSI 1560  
DB 4822 PELHELRLGL-----ECTALS-----GSTDK-----IITTTTITTVTKVISADGKEIVTEQKT 4868  
QY 1561 SLELKRAEPTVNDISELSTGLKHPKDSATTMKMLAAKEL-----DD-----AKP 1606  
DB 4869 VTTTDSSEPDSEKVVVTTTTRTTSERDQLLPKEVALLRGLYRASTPGSEDEDELLLGSP 4928  
QY 1607 AQOLHILQOHFA-----KDVVGDERYEAVERNKLKLVIRQQAADSHSMELSASHSTT 1659  
DB 4929 RSATSYELQHSVGSKRSKRLDADGDE-----SQDDIPPQYGSSEHSTA 4972  
QY 1660 YNNLRINDGIVELLHKHFDAAIPASSAKRUGEMMNNDPALKDILIKLOLSTPF----- 1713  
DB 4973 RSILLPRTADPNA-----TSFYGALPDSF-----DVVMKPSSTEPITIOGAP 5013  
QY 1714 -----SSASVSMELKDGLREOT-----EKAILDGKVGREEVGLFQDRNNLRKYSVSVSQ 1763  
DB 5014 SGDSQSSESSESS-----SQTWAGHKFLDQADKDFORALEE---HVQARGAEVMSVYAKY 5066  
QY 1764 SVSKSEGTWPALLIGTSN-----SAAMSERNIGTINFKYQDQNTPRRFTLEGG 1814  
DB 5067 SYSPSKAEEMEIQVSGTAEQRFPFLSDVQARVAESGFAIVGSAQQOQOQER-----GGE 5122  
QY 1815 IAQANP-----QVASALTDLKKEGLE 1835  
DB 5123 VEQAVPTTAVTATATASSTGALPKDRLE 5153

RESULT 11  
ID Q9NJ17 PRELIMINARY; PRT; 5476 AA.  
AC Q9NJ17;  
DT 01-OCT-2000 (TReMBLrel. 15, Created)  
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)  
DE SPLIT ENDS.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Rebay I., Chen F., Hsiao F., Kolodziej P.A., Kuang B.H., Lavery T.,  
RA Suh C., Voas M., Williams A., Rubin G.M.;  
RT "A Genetic Screen for Novel Components of the Ras/Mitogen-Activated  
RT Protein Kinase Signaling Pathway That Interact With the yan Gene of  
RT Drosophila Identifies Split ends, a New RNA Recognition Motif-  
RT Containing Protein.";  
RL Genetics 154:695-712(2000).  
DR EMBL; AF184612; AAF26299.1; -.  
SQ SEQUENCE 5476 AA; 590531 MW; 93FAA8C7860770C2 CRC64;

Query Match 2.5%; Score 240.5; DB 5; Length 5476;  
Best Local Similarity 19.5%; Pred. No. 0.00023;  
Matches 422; Conservative 256; Mismatches 744; Indels 739; Gaps 106;  
QY 6 LGTEHKAHVHTAAHPVGHGVALQOQSSSSPQNAAS-----LAAEG-----KNRGKMP 55  
DB 25 LVTOQANNSTLAN-----SNSSPSVSASAVFATAAGGSSERSNRDRPY 72  
QY 56 RIHOPSTAADGISAHHQKKSF-----SLRGCLGTGK--FERSAPOGQPGTTHSKGATL 107  
DB 73 RINGSASVQGGGINSNTTTTAACTAGGSGSAGTGTGCLVGGPGVP-----QAL 125  
QY 108 RDLARDGETOHEAA--AP-----DAARLTRSGGVKRRNMDMAGRPYKGGSGEDKVP 160  
DB 126 GDRSSTQNIHQHOSARVAPQOSWYEAATAATTAOLKS-----SGSG----- 168  
QY 161 TQQRHOLNFCQROTMLSMAHPASANAGDRLQHP-----PHI--PGSHHIKE 210  
DB 169 -----NAG-----ASAAVGTMTSSPINHHPHQHPHQPHTYSS 206  
QY 211 EPVG-----STSKATTAHADVEIAQEDDD-----SEFOOLHOORLAR-- 248  
DB 207 PVVGAGCPSAAQGOFOIQSQSOTTAVHRSVAYAGSAADDLLNTATSRNMLLHSSKLNL 266  
QY 249 -----ERENPPQPKLGVPATPISARFQPKLTAVAESVLEGTDTTQSPKPO--- 294  
DB 267 LKGAGATGGGERSGESPPRAGGATPLTTSTITNNSFSSSNLNTITTTATPTMTIAS 326  
QY 295 -----SMLKXSG--AGV-----TPLATVLDKGLQLOAPDNPALNTLLKQT 333  
DB 327 GAAGSVGLGSAEACVCSNGSTASGDILNVAAVLAAVDNG-----VPTPIRTRHNL-- 380  
QY 334 LGKDTQHYLAHHIASSDGSQHLNLDNKGHLFDIKSTATSYVLHNSHPGEIKGLAQAGTG 393  
DB 381 -GRSTTSSSRSHSRSPSY-----SSSHSSSSSSSHSSHS--HASSPVQSSGNCAMA--- 429  
QY 394 SVSVDGKSGKISLGGSTOSHNTKMLSQPGEAHRSLTGTWQHPAGAAARPGESIRLHDDK 453  
DB 430 -----EGRSSR-TVNSVTVTNS--NPSGTA----- 453  
QY 454 IHLHPELGVQWSADKDTHSQLSROADGK-----LYALKDNRTL-----ONLSDNKSS 502  
DB 454 --VTVSSAGVGGCGGSS 511  
QY 503 KLVDKIKSYSDVRGQVA---ILTDTGPHKM-----SIMPSSLDASPEHSITSLSHFA 552





DT 01-OCT-2000 (Tremblrel. 15, Created)  
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
 DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)  
 DE SPLIT ENDS LONG ISOFORM.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Kuang B., Wu S., Shin Y.-A., Luo L., Kolodziej P.;  
 RT "Split ends encodes large nuclear proteins that regulate neuronal cell  
 RT fate and axon extension in the Drosophila embryo."  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF221715; AAF34661.1;  
 SQ SEQUENCE 5534 AA; 599188 MW; 4037E27833D0C622 CRC64;

Query Match 2.5%; Score 240.5; DB 5; Length 5554;  
 Best Local Similarity 19.5%; Pred. No. 0.00024;  
 Matches 422; Conservative 256; Mismatches 744; Indels 739; Gaps 106;

QY 6 LGTEHKAAYHTAAHPVGHGVALQGGSSSSPQNAAS-----LAEG-----NRRGMP 55  
 DB 76 LVTQYANNSTSLAN-----SNSPSSVSASAVFATAAGGSSRSRRDRPY 123  
 QY 56 RIHPSTAAAGISAAHQKSF-----SLRGCLGTRK--PERSAPOGPGTTHSKGATL 107  
 DB 124 RNGSASVQGGGINSSTTTTAACTAGGSGGAGTGTGVLVGGPGGVP-----QAL 176  
 QY 108 ROLLARDGETOHEAA--AP-----DAARLTRSGVKRRNMDMAGRPVMRGSGGDKVP 160  
 DB 177 GDRSSFTQNHQHSARVAPQSWYEAATAATTAQLKS-----SGSG-----219  
 QY 161 TOQRHOLNFGOMRQTMLSKMAHPASANAGDRLOHSP-----PHI--PGSHHEIKE 210  
 DB 220 -----NAG-----ASAAVGFMTSSSPINHHPHOHLQNPQHPHYTSS 257  
 QY 211 EPVG-----STSKATTAAHRAVEIAQEDDD-----SEFQOLHQORLAR- 248  
 DB 258 PVVGAGSCPSAAGCQPOIQSQSQTAVHRVAYAGSADLLNTATSRNMLLHSSKLNKL 317  
 QY 249 -----ERENPPQPKLGVATPISARFQPKLTAVAESVLEGTDTTOSPLKPQ--- 294  
 DB 318 LKGAGATGGGERSGSESPEGAGATLTTTSTINNSFSSNSLNTTITATPTMTIAS 377  
 QY 295 -----SMKSG--AGV-----TPLAYLDKGLQLAPDNPALNTLLKQT 333  
 DB 378 GAAGSYGLSGAEGAGVCSNSGTASGDLNVAAVLAAAVDNG-----VPTHPIRTRNLH-- 431  
 QY 334 LGKDTCHYLAAHASSDGSQHLHLLDNKGHFDIKSTATSYSLVHNSHPGEIKGLAQAGTG 393  
 DB 432 -GSTTSSSSSHSRSPSY-----SSSHSSSSSHSSHS--HASSPVQSSGNCAMA--- 480  
 QY 394 SVSVDGSKRISLGSCTQSHNKTMLSQPEAHRSLLTGIWQHPAGAARPOGESIRLHDDK 453  
 DB 481 -----EGRSSR--TVNSVTVTSS-----NPSGTA-----504  
 QY 454 ITHLPGLVQWQADKDTHTSOLSRQADGK-----LYALKDNRTL-----QNLSDNKSE 502  
 DB 505 --VTVSAGVGGCGSSSSSSSSSSSSGSCGLTANVPVHSEDNRPLAIRVRLPARSSDT 562  
 QY 503 KLVDKTKSVSDRGQVA-----ILTDTPGRHKM-----STMPSLDASPESHISLSLHFA 552  
 DB 563 SLKDLG-LFHEYKKGKVTWKVYVCONSERVALVCFKPKDDVEKALEVSHDKHFFGCKIEV 621  
 QY 553 DAHQGL-----LHKS-----ELEAQSA-----ISHGLVWADSEKGL 586  
 DB 622 EPYQGYDVEDNEFRPYEAELDEYHPKSTRTLFIGNLEKDTITAGELRSH-----FEAFGEI 676  
 QY 587 FSAAIKQK-----DG-----NELKM-----KAMPOHA--LDEHF 614

Db 677 IEIDIKKQGLNAYAFQCYSDIVSVVKAMRKMDGEHLGNRIKLGFKSMPTNCWID--- 733  
 QY 615 GHDDHIS-GFEHDDHGGOLNALVKNFRQOACPLGNDHQFHPGWNLTDLALVDNQLGLRH 673  
 Db 734 GVDEKVESLQSOFTFRGAVTKVSDR-----NRQLALVLDQV--- 773  
 QY 674 TNPEHPEILDMHGLSLALQEGKLHYFDQLTKGWTGAESDC-----KOLKGLDGAAYL 727  
 Db 774 QNAQA-AVKDM--RGTI-LRRKKLV-----DFASRECQDAFYDKQEKQOQSS--- 818  
 QY 728 LKDGKVRKLNINQSTSIKHCTENVEFLPHVRNKE-----PGDALQGLNKDKKA-- 777  
 Db 819 ---GSPNFRFERYESSASSLOSRSRASSFRHQNNSNDCCSPINTPGAGSSGSISSANLIN 875  
 QY 778 QAMAV-----IGVNKYLALTEKGDIRSFOIKPGTQOLERPAQTLRSRIGISGLKDIHVDHK 833  
 Db 876 QSTINISNIGTNACSAM-----PAPSLASAVS-----CNVNAS 910  
 QY 834 QNLYALTHEGEVPHQPREAMONGAESSWHKLALPOSSEKLSLMSHSHKPTATFEDGS 893  
 Db 911 GTVPAST-----SMPSGVSSSS---SSLEPMSPAAL-----AQRHMYVRNARQTV 951  
 QY 894 QHOLKAGWHAYAAPRGPLAVCTGSGTQVFNRL-----MQG-----VKGK 934  
 Db 952 DCFNEVRLRFRSEE---VSGGAGNSTQFEDVRCDSPVTARQGSVAVNCFCTPTAAVGE 1008  
 QY 935 VIPSGSLTVKLSAOTGTMGAERKVSFKSERIRAYAFNPTMTSPRIKNAAYATQHWG 994  
 Db 1009 SIDGTLNNOITGCAEGFTGSGGSLSRRCGK-----TPK-----1044  
 QY 995 QGREGLKPLVEMQALIKOLDAHNVHRNAPQPDLOSKLETLDLGERGAELLNDMKRPRDE 1054  
 Db 1045 ----DLHPVHNQRIQLAEQVE-----ECPSSGDEG--VYSPRKRITKMD 1081  
 QY 1055 LEQSATRSVTVLGQHGVLKNGEINSEFKPSPGKALVGSFNV-----NRSGDLSKSLQ 1109  
 Db 1082 YHHHHHS-----NAGVESTGEHSSINKPSP--LLLSNCDVHIDPLNR-----KSEI 1127  
 QY 1110 QAVHATPPSAESKLSQMLGHFVSA--GVDMSHOKGPIGLQRDPNDKLTALTKSLRLIDT 1167  
 Db 1128 RRVSETSGSPS--IKFPGHLPAPQSLMLSCRRPSIDVGA-----LSALSASSAF--- 1176  
 QY 1168 VTGELHELADKALVSDHKPDADQIKOLRQOFDTLREKRYESNPVKHYTDMGFTHNKAL 1227  
 Db 1177 ----RGI VGASSMDQOHHMNSAAAKRRRVTTTMOQPSSSS-----TTNSS 1220  
 QY 1228 EANYDAVKAFINAFKKEHHGVNLTTRTVLESQGSAAELAKLKNLTLLSDSGESMSFSRSY 1287  
 Db 1221 GSGLGGSISLTPA-DEYHHVSRGRGHQLHSHSHEAS-----GGESADGSRP- 1267  
 QY 1288 GGVSTV-----FVPTLSKKVPVP-----VIPGAGITLDRAYNLFSFTSG 1328  
 Db 1268 --GTPICDERPEVLTPEPRRLPPPRVRERTRDVMWMLPLPKFGLVF---FQQQQRSSG 1322  
 QY 1329 GLNVSFGRDGGVGNINVATGHDVMPYMTGKKTSGNASDWLSAKHKISPDLRIGRAVSG 1388  
 Db 1323 G-----AGNSYLQQLGGSTGGGCG-----IGAASS 1352  
 QY 1389 TLQGTFLNSLKFLTEDELPGFHGLTHGLTLPAAELLQKQIEHMQKQSKLTFSV---D 1444  
 Db 1353 AC--SLNNS-----SLNASGCMSCSGSTFLPSPSSRWRSSSHQONNHOQOQSLHGS 1407  
 QY 1445 TSANLDLRA-----GINLNEGSKPVGTA-----RVNAG-----1474  
 Db 1408 SSSNTCLMASPARPSRLSSNSDSDPVQONAGSPSLDERLNRFEENYERWSSGSSREHI 1467  
 QY 1475 -----LSASANLAAG--SRERSTTSGOFGS-----TTSASNRPFTLN--- 1510  
 Db 1468 SGHTPPSATPQWLSMHMNLSTGLNSHOTSSAGNSNSSGGIVSSSASNSRHFLDDEL 1527  
 QY 1511 -----GVGAGANL--TAALGVAHSSHTHEGK 1533  
 Db 1528 QPSDIVKSVLAKSVFDDDDFORLNKNQWYDPSSSDSFALGSSSNIVTGSSILVANVSRHPGG 1587

QY	1534	PVGIFPFTNTYSAALALDNRSTQSISLELKRAEPTVTSNDISLSTLTKHFKDTSATTK	1593
Db	1588	P-----CSGNTSPALP-----NLAATKATPIIGNCSSGGLNSTGS--KSAGLLO	1629
QY	1594	MLAALKELDDAKPAEQHILQOHFSADKVVGVDEREYAVRNKLKLVIRQQAADSHSMELGS	1653
Db	1630	RLSSLSPWN--SPOASMSPNPSPPSVGG-----VTACLGQLTKPAAPGTASAGL--	1679
QY	1654	ASHSTTTNNLSRINNDGIVELLHKKHFDAAALP-ASSAKRLGEMMNDPALKDIKLOLOSTP	1712
Db	1680	-SGTAASSSPAANSGETKGLQVFPFSPHPPLPMTAPPAVQAPPLPEMGKQSRUTG	1738
QY	1713	FSSA---SVSMELKDG-----LRQTEK-AILDGKVGREEVGVLFQDRNNLRVK-SVSVS	1762
Db	1739	QSSGNLTKLSVDPGQSSPARVOLQKSASVPGSTNVGAPSSLSLSTTASVETSASIS	1798
QY	1763	QSVSKSEGFNTPALP-----LGTNS-AAMSMERNIGTINFEYQODQ	1803
Db	1799	SSTSGNSSLTSAAHVOKPOOSTFVEEHTKSGTSTQSQSSSSSKISSTHKLKLSKH	1858
QY	1804	N 1804	
Db	1859	N 1859	
RESULT 14			
ID	Q9ZHL3	PRELIMINARY; PRT; 4152 AA.	
AC	Q9ZHL3		
DT	01-MAY-1999	(TrEMBLrel. 10, Created)	
DR	01-MAY-1999	(TrEMBLrel. 10, Last sequence update)	
DT	01-JUN-2000	(TrEMBLrel. 14, Last annotation update)	
DE	LARGE SUPERNATANT PROTEIN 1.		
GN	LSPAL		
OS	Haemophilus ducreyi.		
OC	Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;		
CC	Haemophilus.		
OX	NCBI_TaxID=730;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=35000;		
RX	MEDLINE=99030326; PubMed=9811662;		
RA	Ward C.K., Lumley S.R., Latimer J.L., Cope L.D., Hansen E.J.;		
RT	"Haemophilus ducreyi" secretes a filamentous hemagglutinin-like		
RT	protein.;		
RL	J. Bacteriol. 180:6013-6022(1998).		
DR	EMBL: AF057695; AAC79757.1;		
SQ	SEQUENCE 4152 AA; 456179 MW; 7D82DEDC988AB8F3 CRC64;		
Query Match 2.5%; Score 237.5; DB 2; Length 4152;			
Best Local Similarity 18.9%; Pred. No. 0.00022;			
Matches 345; Conservative 244; Mismatches 636; Indels 599; Gaps 88;			
QY	350	GSQHLLDNKGHLEDFDKTSTSVYLHNSHP-----GEIKGKLAQAGTGSVSV	397
Db	835	GGGGLNLAKGNTN-DSNSTALVHLNSNDINLANNKVNIYEIYSQ-----AGNISV	888
QY	398	DGK-----SGKISLGSQTSHNKTMLSQPGEAHRSLTGLWQHYPAGAARPOGESIRL	449
Db	889	EAKLLHNDVKLSGNTT-TTKSGNATVKT-----NSIGGLHDANSIRV	931
QY	450	HDDKTHILPELVQWQADKTHSLSQADGKLYALKDNRTLNLSNDKSSSEKLVYDKIK	509
Db	932	GELTN-----GKAFDL--NQLVALRGKIYA-GSNLTPK-----AKEGEKOKSTA	976
QY	510	SYSDVORGVOAILTDPGRHKMSIMPDLASPESHI-SLSLHPADAHQGLLHGKSELEAQ	568
Db	977	OAKIIRGNTINV-----KNKLEYGSNDV--ENNMRSMQVNL---YEKIFNGDNPI---	1022
QY	569	SVAISHGRVLVADSEKGLFSAAITPKQGDGNGELKMKAMPQHALDEHF-----GHDHQSIF	623

Db	1023	TLTLKNGVTFPAKDFSNRRRRAS--NDGEGTNKKTDFDNVAHLIEAFSCSYNGNDHRAS--	1078
QY	624	FHDDHGOLN-----ALVKNFRQOHACPLGNDHQPQWN-LT-----DALVID--NQ	668
Db	1079	--DDGHVKSPLYLLVLAQAVNTEGENYKLTALQHIFGPNWDLTTTNNDDTTINDKNWQ	1135
QY	669	LGL-----HHTNPEPH--EILDMGHLSLALOG-----KL	697
Db	1136	LKLKWEFKKNGENNNHNSINLNIYPADEGVERAKIFAGVLRNGTNGVEDKVVQELNDKAKK	1195
QY	698	HYFDOLTKGMTGAESDCKQLKKG-LDGAAYLLKDG-----VKRLNINOS	741
Db	1196	EYEDKFAKKFQGRFK--SRFQNGEFDWAGDMAKEGNESYSGKETEEKYNGIKKEHTVNIG	1253
QY	742	TSSIKHGT---ENFSLPHVRNKPEPGDALOGLNK-----GIDKSIISELAAQYIYAKADVPDPRV	1307
Db	1254	KHEIKVPTVSFENLNINHOQDKSD-----GIDKSIISELAAQYIYAKADVPDPRV	1307
QY	774	--DDKA-----QAMAVIGVKNYLAL-----TEKGDIRSPQIKPGTOOLERPQOT	815
Db	1308	AONDKAVDEGLYRTRLSYINQNNYLKAKYFNQLDTEDDKLKGIK-RIGDNYFEH--QL	1364
QY	816	LSREGISGELKDIHVDHKONLY--ALTHE-----GEVPHQPREAWONGA	857
Db	1365	ITR--LIEKVADNHLTLKHLGLHDLALVKKLIDSASIQAKDLNLKVGCEAL--TKEQKDNLK	1420
QY	858	ESSWH-----KLALPOSESKLSDMSHEHKPIATFEDGSHQKAGGWHAYAAPE	909
Db	1421	EDIVWVYVTEVNAQEVLPQVYLAKOTIEVEKQRCVGT-----QOIRAGIIDVKVDV	1474
QY	910	RGP-----LAVGTSGSQTVFN-----RLMQGVKGVIPGSGLTVKLSAQTGGMTGABG	957
Db	1475	RNTGTIAGVAVGLEAKNKLKNTGDIILSQLSLKLVGKKGLESTGVT--YVDETG--ATKV	1529
QY	958	RKVSSEFSPRTIRAYAPNPTMTSPRIKNAAYATOHWCQREGKLPLEYMQGALIKOLDAAH	1017
Db	1530	KRARIKSEGI-----YLETDKD-KNYD-LTASELKGNTG-----QIKAK	1567
QY	1018	NVRHN-APODLQSKLETLDLGEHGAELNDMKRFR-----DELOSATRSVTV	1065
Db	1568	DLNLNDIYTSYKYKYEKL-FGKNGEIGDRVITQTSQAKSVGTDAFDFHLHLSLEGDVNQ	1626
QY	1066	LQOHQGVLSKNGEINSEFKPSGKAL-----VQSFNVNRSQGDLSKSL-QOAV	1112
Db	1627	TGSLNLANRTTGVVKGDFNTKAGKDLFHRQIDTPTVTSYASASGGGSGAGLSLTDQGV	1686
QY	1113	HA-TPSAESKLOSLMGHEV---SAGVDMSHOKGEIPLGRQDRDPNDKTKALSRILDT	1167
Db	1687	ETVYNTKATAGANADVTFNFKRTRETETSLTHRNEF-----NALSGELIYVMGK	1735
QY	1168	VTIGEL-----HELADKAKLV-----SDHKPDA	1190
Db	1736	ADIGGVDINRDVEVIKTPPEIAAEQAAEAKAEVKEANEASEATAKETEFENDNVAEK	1795
QY	1191	DQIK-----QLRQOEDTLREKRYESNPVKHYTD--MGFT-----HNKA--LEA	1229
Db	1796	DKTKPKFKLTDDEIAAAEFETKGEDFFAAYKAREEDRKGFLLSAFOIESTKARDEKET	1855
QY	1230	NYDAVKAFINAPKKEH---HGVNLTTRTVLESQGS-----A	1262
Db	1856	TYVELKGVGCAEAEASAAADAIISNKAQIIDTQNGLKQDGTVALQEAQSDVLNLTADLA	1915
QY	1263	ELAKKLKNTLLSLD-SGESMSFSRSYGGG-----VSTVFVPTLSKKYPVP	1306
Db	1916	GASAKLLFELSTIEKKSRSASDGRSILGRLNLAARGDITLNNVETTENSJLSKARDN	1975
QY	1307	VIFAGIT--LDRAYNLSFSRTSG---GLNV-----SFGRGDGGVSGNIMVATGHDVPMY	1356
Db	1976	VYNSGVTEQKDESNQSOLKVTAGASSGCGVMAGCSAGVSAGVSGS-----YN	2024
QY	1357	TGKKTSGAGNASDMLSAKHKISPDIRIGAAYSGTLQGTQNSLKFKLFEDELPGFIHGLTH	1416
Db	2025	ESNTESTSHTNLSLLRGK-----SLRVEA-----GKDFNLISSNVQDVLHLDVKGDTN	2072



QY	1417	GTLT	PAELLOK--GIEHOMKQGSKLTFTSYDSTANLDRAGINLNEDGSK----	PNGVTAR	1470
Db	2073	VYSKQDSY	SKRGGVNVYSAGVGVSTAGARPNGSVGLGSAENENSKIVKQOAGISAK	2132	
QY	1471	VSAGLS	SASANLAAG-----SRERSTTSQGFSTTSASNRPFTF	1508	
Db	2133	RITGEIN	LNILNLTGGYIENKGNPDELNVKGDITTHELKDEHHKDGSGFSGVSGVSETGV	2192	
QY	1509	LNGVGAG	ANLTAALGVAHSS---THEKGPVGIFPAFTSTNVSAALALONRTSQSISLEL	1564	
Db	2193	VNVNGR	VEQKHYEATQHSISGIWTKGTGNF-----KTRDSOSTEV	2236	
QY	1565	KRAEPVTS-----NDISELSTLIGKHFKDSATTMKAALAKELDDAKPAEQ	HLILQQHF	1617	
Db	2237	HRDDT	IAATNFNFELGDIAELAKK--GKEKWDNRSAKTTS-----	2274	
QY	1618	SAKDVVG	DERYEAVRNKLKLVIRQAA-----DSHSMELGSAHSHTTYNNLSRINNDGI	1671	
Db	2275	SSQSAH	DPGRSVENGYSYELPRFKTADNDAGVDSPRIKGEAQOATL--ALTKAGNDVI	2333	
QY	1672	VELLKH	FDAALPASSAKRLEGMNNDPALKDIIKLOLSTPESSASVSMELKDGLREQTE	1731	
Db	2334	-----	FEVQSLTKARQPSVLVDSPPAEIETPALTRPQVK	2366	
QY	1732	KAILDGK	VGREEVGVLFODRNLNRKVSYSVQ---SVSKSEGFNTPALLGTSNSAAMSM	1788	
Db	2367	SNI-----	AESIEVPQFRTKVSDGDGEGNVAETTFPT--NKAATSS	2404	
QY	1789	ERNIGT	INFKYQDQONTPRRFTLE	1812	
Db	2405	TODVG-----	DTPTPRALRLE	2420	
RESULT	15				
ID	097205	PRELIMINARY;	PRF;	3012	AA.
AC	097205;				
DT	01-MAY-1999	(TREMBLrel. 10, Created)			
DT	01-MAY-1999	(TREMBLrel. 10, Last sequence update)			
DE	01-JUN-2000	(TREMBLrel. 14, Last annotation update)			
DE	L2969.6	PROTEIN.			
GS	L2969.6.				
ON	Leishmania	major.			
OC	Eukaryota;	Euglenozoa;	Kinetoplastida;	Trypanosomatidae;	Leishmania.
OC	NCBI_TaxID=5664;				
RP	[1]				
RN	SEQUENCE FROM N.A.				
RA	Oliver K., Quail M., Lawson D., Harris D., Rajandream M., Ivens A.,				
RA	Barrell B.;				
RL	Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AL034357; CAB39097.1;				
DR	INTERPRO: IPR001201;				
DR	INTERPRO: IPR002016;				
DR	INTERPRO: IPR002934;				
DR	PFAM: PF01909; NTP_transf_2;				
DR	PROSITE; PS00435; PEROXIDASE.1;				
QY	SEQUENCE	3012	AA: 309440	MW: 719A0D5320D648C4	CRC64: 1.

[illegible]



Db 1337 -----DDDEFARLASSGTV-----PGAGATS--TCNKSGVGHTDSASA 1373  
QY 1237 FINAFKKEHHGYNLT---RTVLESQSAELAKKLKNTLLSLDSESMFSRSYGGVST 1293  
Db 1374 -----SRLHNPDAETAEDTEYGTGVDVEQAEVEFRGQIEKD---YNFSTSV--DLSA 1421  
QY 1294 VFVPTLSKKVPVPVTPGAGITLDRAYNLSFRSRTSGGLNVSFGRDGGVSGNIMVATGHDVM 1353  
Db 1422 LFAPSWGSGMP-PVAPPVA-----AASGLSI-----GGASSS---SSGLALQ 1459  
QY 1354 PYMTCKKTSAGNASDWLSAKHKISPDLRTGAAVSGT--LQGTLONSLKFALT-----ED 1405  
Db 1460 TATTRLHNSASLSSP-----RESAASGSADLKGT-PRSLPPAMTSGLHTDED 1506  
QY 1406 ELPGFTHGLTHGLTPAEILLQKIEHQKQSKLTSVDTISAN-LDLRA-GINLNEDG-S 1462  
Db 1507 NIAAHAQPLPRSKAAPAD-----ADMSANEAEELRCQVRDLADGGAA 1547  
QY 1463 KPNGVTA-----RVSAGISASANL-----AAGSRERSTTSQGFGS 1497  
Db 1548 EPEGRTASGGDAAAEEAAEAHAGTETVADEAATAQLSPVPTAAASGERAAV-QTAA 1606  
QY 1498 TTSASNRRPTFLNGVAGANLTAALGAHSS---THEGKPVGIFPAFTSTNVSAALADN 1554  
Db 1607 APPASODAVT-RNSIVLGATDGRGCGHLYYVPVHEG-PLFFVQSITATRPVPIKLTDK 1654  
QY 1555 RTSQSISLELKRAEPTVSNDSIELTSLGCKHKFKDSAT-----TKMLAALKELDDAKP 1606  
Db 1665 ATGTVKVDITFAGEHWR---MOLTKSLLEVFPFHARTLILFLKYCVRSLSAGESEPGGV 1721  
QY 1607 ABQLHILQGHFSAKDVGVDERVEAVRNKLVIRQQAADSHSME-----LGSASHS 1657  
Db 1722 SFAYILMVLHF-----YNECQRILLRLRERNAASSSAEAOGRQAGADAGEHQH 1771  
QY 1658 TTYNLSRINDGIVELLKHHPDAALPASSAKRLGEMMND-----PALKDIIKQLQST 1711  
Db 1772 -----HQLOKDRAAEEALRRGSAADNSAASPAPRLAAVYQDASTT 1813  
QY 1712 ----PSSASVSMELKDLREQTEK-----ATLDGKVGREEVGVLFQDRNNL 1754  
Db 1814 AGPAPLTLGLLEYMLGEYLARVEORHAQLVEEAGARRSALDA-AGAGDVAELYSSESGNI 1872  
QY 1755 RVKSVSVSQSVKSEGFNTPALLLGTSNSAAMSMERNIGTINPKYGOD 1802  
Db 1873 Q-KTV-----PA--LGRASSSTAVATGACATIKAAAYSD 1903

Search completed: June 5, 2001, 18:23:10  
Job time: 391 sec







CC CONTACT AND, IN LYMPHOCYTES, AT CELL-CELL CONTACTS (BY  
CC SIMILARITY).  
CC -!- SUBUNIT: BINDS WITH HIGH AFFINITY TO VINCULIN AND WITH LOW  
CC AFFINITY TO INTEGRINS (BY SIMILARITY).  
CC -!- SIMILARITY: CONTAINS A DOMAIN FOUND IN BAND 4.1, EZRIN, MOESIN,  
CC RADIXIN, AND TALIN.  
CC -!- SIMILARITY: IN THE C-TERMINUS TO YEAST SLA2 AND C.ELEGANS ZK370.3.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: AF078828; AAD13152.1; -;  
CC EMBL: AF177198; AAF23322.1; -;  
CC MIM: 186745; -;  
CC InterPro: IPR00159; -;  
CC InterPro: IPR00259; -;  
CC InterPro: IPR00258; -;  
CC Pfam: PF00373; Band\_41; 1.  
CC Pfam: PF01608; I\_LWEG; 1.  
CC PROSITE: PS00660; BAND\_41\_1; 1.  
CC PROSITE: PS00661; BAND\_41\_2; 1.  
CC PROSITE: PS00657; BAND\_41\_3; 1.  
CC Structural protein: Cytoskeleton.  
CC DOMAIN 173 332 BAND 4.1-LIKE.  
CC CONFLICT 1227 1227 L -> S (IN REF. 2).  
CC CONFLICT 1966 1966 R -> A (IN REF. 2).  
CC SEQUENCE 2541 AA; 269827 MW; 0FEDA1A9E16EE780 CRC64;

Query Match 2.4%; Score 227.5; DB 1; Length 2541;  
Best Local Similarity 19.5%; Pred. No. 0.0058;

Matches 449; Conservative 297; Mismatches 846; Indels 707; Gaps 114;

QY 11 KAAVTHAHPVG---HG-VALQ---QGSSSSPQN-----AAASLAAGKNGRMR 56  
DB 428 KSTVLQQYNRVKGVEHGSVALPAMRSGAGSPENFGVSMPPAQOQITSGQMRGHMP 487  
QY 57 IHOPSTAADG-----ISAHQKKSFLSLRCLG---TKFERSAPOGQGTTHSK- 103  
DB 488 LTSQAQALGTINSSMQAVQAQATLDDFTLPLGLQDAASKAWRKNKMDKESHEHSQV 547  
QY 104 -----GATRLDLARDGGETOHE-----AAAPDAALTR--SGGVK--RRNMDM--AGRP 148  
DB 548 DAITAGTASVNLTAGDPAETDYTAGCAVTTISSNLTEMSRGVKKLLAALLEDEGGSGRP 607  
QY 149 MYKGGSGEDKVPQQRHQLNPNFGQMRQMLSKMAHPASANAGDRLOH-----SPPHIG 203  
DB 608 LLQAAGLAGAVSELLRSQAASAEPRQLLQAGNVGQA--SGELLQOIGESDTPHFQD 666  
QY 204 SHHEIKKEPVGTSKATTAHADRVETAQEDDSEFOOLHOORLARENPQPKLGVA 263  
DB 667 ALUQL-ARAVAGAAALVUKAK--SVAQRTEDSLGQ---TQVTAATQCALSTLSOLVACT 720  
QY 264 PLSARFQPKLTA-----VAESVLEGTDTTQSPKPKQSMKKGSGAGVTPLA 310  
DB 721 KVVV---PTISSPVQEQELVEAGRLVAKAVEGCVSAQATQEDGQLLRGVGAAT--AVT 775  
QY 311 LDKGLQLAPDNPALNTLLKQTKDQTHYLAHHASSDGSOHLLDNKGHLFDIKSTAT 370  
DB 776 -----QALNELL-----QHVRAHATGAGPAGR-----YD-QATDT 804  
QY 371 SYSVLHN-----SHPEIKGK---LAOACT---GSVSDGKSGKISLGSTQSHNKTMLS 419  
DB 805 ILTVTFSSMGDAGEVMGQARILAAQATSLDLVNAIKADAE-----GESDLENSKLLS 858  
QY 420 OP---GEAHRSLLTGTWHPAGAAARPQGESIRLHDDKIHLHPGLGWQASDKDTHSLS 476  
DB 859 AAKILLADATKAVV-----EAAKGA-----AHPD-----SEEQOORLR 891

QY 477 ROADCKLYAKDNRTLQNLSDNKSSEKLVDKIKSYSDVQRGQVAILTDTFGRHKMSIMPS 536  
DB 892 EAAEG--LRMATNAAQ-----NAIKKKLVORLE-HAAKQAAASATQITAAQAHAAS-TPK 943  
QY 537 LDASPESHLSLIH-PADAHQGLLHGKSELEAQSVALSHGRLVVADSE-----GKLPSA 589  
DB 944 ASAGFQPLLVSQSKAVAEQIPLLVQGVRSQAQPDSPSAQIALTAASQSLQPGKMWAA 1003  
QY 590 A---TPKQDGNELKMKAMPQHALDEHFGHDHOISGFFHDDHGGOLNALVKN----- 637  
DB 1004 AKASVPTIQD---QASAM-----QLSQCAKNLGTALAE 1035  
QY 638 ---NFRQOHAC-PLGNDHQHPGWNL-----TDALVIDNOLGLLHTNPEHEIL-----D 683  
DB 1036 TAAQRAQEAACGPLEMDSALSVVONLEKLOEVKAAARDGKL-----KPLGETMEKCTQD 1090  
QY 684 MGH-----LGSIALQEGKLYHYFDOLTCKGWTGAESDCKOLKKGLDGAAYILKDG 731  
DB 1091 LGNSTKAVSSAIAQLLGEVA--QGNENYAGIAARDVAGG--LRSLAAARGVAALISDP 1145  
QY 732 EVKRLNINOSTSIIKHGTENVFLPHVRNKPDPGDALQGLNKDDKAQAMAVIGVKNYL-- 789  
DB 1146 AVQAIVLDTASDVLDKASSLIEEAKKAAGHPGPESQORLAQVAKAVTQA---LNRVCSC 1202  
QY 790 -----ALTEKGD---IRSFQIKPGTQOLERPAOTLSREGISG-----ELKDIHV 830  
DB 1203 LPGAQDVNMLRAVGADASKRLLSDLLPSTGTGTFQE--AOSRLNEAAAGLNAQATELVOASR 1261  
QY 831 DHKQNL-YALTHEGEVHFQPREAWQNGAESSSWHKLLALPOSE-----SKLKSIDMSHEH 883  
DB 1262 GTPODLARASRGFGQDFSTFLEA--GVEMAG---QAPSQEDRAQVVVNLKGLISMSSSK 1314  
QY 884 -----KPTATFEDGSOHQKAGGWHAYAAPERGLAVGTSGSOTV----- 923  
DB 1315 LLLAAKALST--DPAAPNLKS-----QLAAAAAR---AVTDSINOLITMCTQAPGQKECDN 1365  
QY 924 -----FNRLMQGVKGVIPGSGLT-VKLSAOTG-----GMT 953  
DB 1366 ALRELETVRELLNPVQIPINDMSYFCCLDSVMENSKVLGEAMTGISONAKNGLNPEFGDA 1425  
QY 954 GAEGKVKSKFSERIRAYAFNPTMTSPRIKNAAYATQHGQOQREGKPLIYEMOGALIKQ 1013  
DB 1426 ISTASKALCGFTEAAQAQRAYLVGVSDP-----NSAQOQGLVEPTQPARA--- 1470  
QY 1014 LDAHVNRNAPQDQLQSKLETLDLGEHG---AELLNDMKFRFEDEQSATRSVTVLGQHQ 1070  
DB 1471 -----NOA IQMACQS-----LGEPGCTQAQVLS-----AATIVAKHT 1502  
QY 1071 GVLKSNGEINS--EFPKSPGKALVQSF--NVNRSQDLSKSLQ-----QAVHAT 1115  
DB 1503 SALCNSCRLASARTTNPATKRFQVQSAKEVANSTANLVKTIKALDGPFTTEENRAOCRAAT 1562  
QY 1116 PPSAE-----SKLQSMIGH-----FVSAGVDMHQKGEIPLGR--- 1148  
DB 1563 APLLEAVDNLSAFANPEFSSIPAQISPEGRAAMEP I VISAQTMLESAGGLIOTARALAV 1622  
QY 1149 -ORPNDKALT-KSRLILDV-----TIGELH-ELA-----DKAKLVS 1184  
DB 1623 NRPDPSPSVLAGHSRTVSDSIKKLITSMRDKAPGOLECEATAALNSCLRDUDQASLAA 1682  
QY 1185 -----DH-----KPDADO-----IKLQRFQFD--T 1202  
DB 1683 VSQQLAPREGISQEAHLTEMLTAVOEISHLIEPLAHAARAEASOLGHKYSQMAQYEPFLT 1742  
QY 1203 LREKRYESNPVKHYTDMGFTH-----NKALEANDVAKAFINAFKKEHHGVNLTTRTVL 1256  
DB 1743 LAAYGAASKTTLSPQOAMLLDQTKTLAELSALQLLYTAKEAGGNPKQAAH-----TQEAL 1796  
QY 1257 ES---QGSALAKKLNTLLSLDSGSMFSRSYGGVSTVFVFTLSKKYVPVPIPGAIT 1314  
DB 1797 EEAVOMTEAVEDELTTL-----NEAASAAAGVVGWVDSI-TQAINQLDDEGMPGEPSGF 1850

QY 1315 LDRAYNLSPSTSGGLNVSGRGGVSGNIMVATGHDVMPYMTG----- 1358  
 Db 1851 VD--YQTTWRTAKATAVTOE-----WTKSNTSPEELPLANQLTSDYXRLASE 1899  
 QY 1359 ---KTSAGNASDWLSAKHKISPDRLRIGAVSGTLQGTQNSLKFKITEDLPFGIHLT 1415  
 Db 1900 AKPAVAANEENEIGSHIKRVQ--ELHGCAALVTKAGALQSPSDAYTKKEL----- 1950  
 QY 1416 HGTLPAPAEELQKIEH---OMKOGSKLTFSVDTSAN-----LDLRAGI-----NLNED 1460  
 Db 1951 ---IECARRYSEKSVHLRALQAGNRGTQACITAAASAVSGIADLDTTIPATAGTLNRE 2007  
 QY 1461 GSKP-----NGV--TARVSA--GLSASANLAAGSRERSTTSQ-----FGST 1498  
 Db 2008 GYETTFADHREGILKTAIVLVEDTKVLVONAAQSGQEKIAQAQSSVAITRLADVVKICAA 2067  
 QY 1499 TSASNNRPTFLGVGAGANLTAALGAHVSSTH--EGKPVGIFPAFTSTNVSAAALADNRT 1556  
 Db 2068 SLGAEDPETQVVLINAVKDVAKALGLDISATKAAAGK--VGDDPAVWQKNSAKVMVTNVT 2126  
 QY 1557 S-----QSISLEL-----KRAEPTVSNDSIELSTLTKHFF 1586  
 Db 2127 SLLKTVKAVEDEATKTRALEATEHRIQELAVPCSPPEPAKTSTPEDFIRMTKGI----- 2182  
 QY 1587 KDSATTKMLA-----ALKELDDAKPAEQLHILQHFSAKDVVGVDERYEAVRNK 1635  
 Db 2183 -TWATAKAVAAGNSCRQEDVIATANLSRRALADMRLCAKAAHYHPEVAPDVRLRALHYGR 2241  
 QY 1636 KLV-----IRQAADSHMELASASH-----S 1657  
 Db 2242 ECANGYLELHDHVLTLQKPSPELKOQLTGHSKRVAGSVTELIOAAEAMKTEWVDPDP 2301  
 QY 1658 TTYNNLSRINNDGIVELLHKEHDAALPASSAKRLGEMNNNDPALDKIILQKQSTPFFSAS 1717  
 Db 2302 TVIAENELGAAAEIAAAKLEQLKPRAPKEADESLNPEQILEAKSI-----AAA 2355  
 QY 1718 VSMELKDLREOTEKAILDGVKEEVEGVILFQDRNNLRKVSVSQSVSKS----- 1768  
 Db 2356 TSALVKAASAAQRE--LVAQCKVGAIPANAL--DGOVWSQGLISAARMVAAAATNNLCEAAN 2412  
 QY 1769 ---EGNTALLGTSNSAAMSMERNIGTINFKYQDQNTPRRTLEG--GIAQAN----- 1819  
 Db 2413 AAVQGHASQEKLTSSAKQVRAASTAQLLVACKVADQDSEAMKRLQAAGNAVKRASDNLVK 2472  
 QY 1820 -PQVASALTDLKKEGLEMK 1837  
 Db 2473 AAKAAAFEEQENETVVVK 2491  
 RESULT 3  
 HLVA\_SERMA  
 ID HLVA\_SERMA STANDARD; PRT; 1608 AA.  
 AC P15320;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 01-NOV-1990 (Rel. 16, Last annotation update)  
 DE HLAYSIN PRECURSOR.  
 GN SHLA.  
 OS Serratia marcescens.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Serratia.  
 OX NCBI\_TaxID=615;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 31-40.  
 RC STRAIN=SN8;  
 RA MEDLINE=88257037; PubMed=3290200;  
 RA Poole K., Schiebel E., Braun V.;  
 RT "Molecular characterization of the hemolysin determinant of Serratia  
 marcescens.";  
 RT J. Bacteriol. 170:3177-3188(1988).  
 RL J. Bacteriol. 170:3177-3188(1988).  
 CC -!- FUNCTION: BACTERIAL HEMOLYSINS ARE EXOTOXINS THAT ATTACK BLOOD  
 CC CELL MEMBRANES AND CAUSE CELL RUPTURE BY MECHANISMS NOT CLEARLY  
 CC DEFINED.

CC -!- FUNCTION: CELL-BOUND HEMOLYSIN, WHICH RELEASES HEME-IRON FROM  
 CC ERYTHROCYTES BY INTERACTION WITH THE ERYTHROCYTE MEMBRANE. SHLA  
 CC REQUIRES SHLB FUNCTION.  
 CC -!- SUBCELLULAR LOCATION: OUTER MEMBRANE.  
 CC -!- SIMILARITY: TO P.MIRABILIS HEMOLYSIN (HPMA).  
 CC  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC EMBL; M22618; AAA50323.1; -  
 CC PIR; A28182; A28182.  
 KW Hemolysis; Toxin; Outer membrane; Signal.  
 FT SIGNAL 1 30  
 FT CHAIN 31 1608 HEMOLYSIN.  
 SQ SEQUENCE 1608 AA; 165078 MW; D669B476FE7DAD51 CRC64;  
 Query Match 2.3%; Score 221; DB 1; Length 1608;  
 Best Local Similarity 18.6%; Pred. No. 0.0059;  
 Matches 372; Conservative 235; Mismatches 689; Indels 700; Gaps 89;  
 QY 33 SSSSPQNAASLAAGKNGKMPRIHQPSTAA-----DGISAHQOKKFSLSR--GC 82  
 Db 23 AASAGAVAAEIVAANGANG---PGVSTAATGAQVVDIVAPNGSLHNGVQDFNVOPGA 79  
 QY 83 L--GTKFSPAPQPGQGTTHSKG-----ATLRDLARDG-----ETQHEAAA 124  
 Db 80 VLNSREAGLSQALQGLGANPLNGREASVILNEVIGRNPSSLHGQOEIFGMAADYVLAN 139  
 QY 125 PDAARLTRSGVGRNRMDMAGRMVKGSGEDKVPYQKRLQNNFGQMRQMTLSKMAH 184  
 Db 140 PNGISQSCGFINTSHSLVGNPLVENVLOG-----YSTFGNRTLSLN--- 185  
 QY 185 PASANAGDRLOHSPHPHPSHHEKEEVPVSGTAKTATHADRV--IAQEDDDSEFOQLHQ 243  
 Db 186 -GTLNAGVGLDLIAPKIDSRGEVIVQDFKQSGKVTSAINAISGLNRVARDGTVAOQ 244  
 QY 244 QRLARERE-----NPPQPKLVATPTISARFQPKLTAV-----AESV 280  
 Db 245 MPTALDSYLLGSMQAGRINIINTAOGSGVKLAGSLNAGDELKVKAYDIRSEVDDASSN 304  
 QY 281 LEGDITDQ-----SPLKPSQMLKG-----SGAGVTPPLAVTLDKG 314  
 Db 305 KNGGDNYQNYRGYVNDRSSQTLTRTELKGNISLVADNHAHLTATDIRGEDITLQGG 364  
 QY 315 KLQIAPDNPPALNTLLKQTLGKDTQHYLAHASSDGSQHLHLLDNKGHLFDIKSTATSYV 374  
 Db 365 KLTL-----DQOQ--LKQQTQHTDDRWYFQWYDV 392  
 QY 375 LHNHPGEIKGLAQAGTGSVVDCKSGKIS-----LGSGTQSHNKMTLSOPGEAH-- 425  
 Db 393 TRE-----REQLQOAGS--TVAASGAKLISTQEDVKLLGANVSADRLSKAARDVHLA 445  
 QY 426 -----RSLLTGTWQH-----PAGAARPOGE-----SIRLHDDK 453  
 Db 446 GLVEKDKSSERGQYGRNHTSSLRTGRWSNDESELSKASELSRSEGLTLKAGRVSTQAG 505  
 QY 454 IHLHPELGWVQWQADKPTHSQLROADGKLYALKDNRTLQ---NLSDNKSSEKLVLDKIS 510  
 Db 506 VHAQRD-----LTIDADNOIQGVQKRTANAKAVRDDKTSWGGTGGGDKNNNS----- 552  
 QY 511 YSVDQRGQVAILLTDPGRHKMSIMPDLASPESHISLSLHFADAHQGLLHGKSELEAQ-- 568  
 Db 553 -----NRREISHASELTSGGTLRL-----NGQQGVTTITSGKARGOKG 589  
 QY 569 -SVAISHGRVLVADSEGLKFLSAIPKQGDGKLNELKMAKMPQHALDEHFGHDIQISGFFHDD 627  
 Db 590 GEVATATHGGLRIDNA-----LSTTVDK-----IDARTGTAFNITSSSHK-- 628

QY 628 HGOLNALVKNFRQOQHACPLGNDHQHPGWNLT-----DALVIDNOLGLHHTNPEPHEIL 682  
Db 629 -----ADNSYSSSTASELKSD-----TNLTLSVSHKDAVDIGSQA----- 663  
QY 683 DMGHLGSLALQEGKHLFYDQLTGKWTGAESDCKQLKGLDGAAYLLKLDGEVKRLNINQST 742  
Db 664 -----SGGE-----LSVESKT 674  
QY 743 SSIKHGTENVFSLPHVRNKPEDALQGLNKKDKKAQMAVIGNVKNYALTEKGD----- 796  
Db 675 GNI-----NVKA-----AERQONIDEQKALTATVNGYAK-----EAGDKQYRAG 712  
QY 797 --IRSFQIKPGTQOLRPAQTLSREGISGELKDIDHVKHQNLYALTHTHEGEVFEHPREAWQ 854  
Db 713 LRIEHTROSEKTRTSENSASSLS--GGSVKLG-----AEKDVTFSGSKLVAD 757  
QY 855 NGAESSWHKLALPQSESKLSLMDSHSHEHKPIATFEDGSOHLKAGGWHAYAAPERGPLA 914  
Db 758 KGDASVSGNKVSFLAADDKTAS-----NTEQTKIGGGFY----- 792  
QY 915 VGTSGSQTVFNRLMOQVKVPIPGSLGVKLSAQTGKMTGAEGKVKSSKESERIRA---Y 971  
Db 793 --TGG-----IDKLGSGVE-----AGYENKTAQOSKAITSGSDV 826  
QY 972 AFNPTMTSTPRPIKNAAYATQHGWOGRGLKPLYEMOGALIKQLDAHNVRHNAPOPDLQSK 1031  
Db 827 KGNLTINARDKL-----TQOGAQ-----HSVGGAY--QENAGVDHLAAADTASTT 870  
QY 1032 LETLDLGEH--GAELLNDMKRFRDELOSATRSVTVLQGHOGVLKNGEINSEFKPSPGKA 1090  
Db 871 TTKTDVGVNIGANV--DYSAVTRPVERAVGKAAKL--DATGVINDIGGIGA--PNVG-- 921  
QY 1091 LVQSFNVNRSQDLSK--SILOQAVHATPPSAESKLOSLMGLHFYSAGVDMSHOKGEIPLGR 1148  
Db 922 ----LDIGAQSSEKSRSSSQAVSS-----VOAGSIDINAKGEV----- 958  
QY 1149 QRDPNKDTALTAKSLRLDVTITGELHELADKAKLVSDHKPDQDQIKOLROQFDTLREKRY 1208  
Db 959 -RDQGTQYQASK-----GAVNLTD-----SHRSEAAANRQDEQSDRT-----R 996  
QY 1209 ESNPKVHYTDMG-----FTHNKALEANYDVKAFINAFKE--HIGVN 1249  
Db 997 GSAGVRVYTTTGSDLTVDAKGGGTQRSNASSAQVTSIDAANGINNVNKKDAIYQCTA 1056  
QY 1250 LP---TRTVLESQGSAL-----AKLKNLTLLSDSG-----ESMSFSRSYGGV--- 1291  
Db 1057 LNGGRKTAVNAGGDIRLDQASDKQSESRSGFNWKAASAKGGFTADSKNFGAGFGGTHNG 1116  
QY 1292 ----STVFVPTLSKKVPVPIPGAGITLDRAYNLSPRTSGGLNVSPG----- 1335  
Db 1117 ESSSSTAQVGNISGQGVKAGRLDTLQ-----TDVKSGQDVSLSAGNKVALQAESTQ 1172  
QY 1336 --RDGVSNCINAVTGHVDPYMTGKKTSGNASD-----WLS 1371  
Db 1173 TRKESKLSGNDLGAQSSDSKSKTGNLNSAGGAFDIKAKVNESATERQGATIASDGKVTLS 1232  
QY 1372 AKHKISPDRLI--GAAVSG---TLQ-----CTLQNSLKFKLTEDELPGFIHCLTHGLTPAE 1423  
Db 1233 ANGKDDALHLQCAKVSQGSAALEAKNGGILLESNAKEQHKN----- 1275  
QY 1424 LLOKGIEHMQKOGSKLTFSVDTSAMLD-----LRAGINLNDGSKPN--GVTA 1469  
Db 1276 -WSLGIKANAKGGQ--TFNKDAGGVDPNTPGTGKTHTLGAGLVGVFOQDKTHANTGIITA 1332  
QY 1470 -RVSNAGLSANLAAGSRERSTTSQFGSTTSASNNRPTFLNGV-----GAGA 1516  
Db 1333 GCVTLNSGKDTFLAGARVDADSVQGVGDHLV--ESRKDVENGKVDVDAGLSHSDNPGS 1391  
QY 1517 NLTAALGVAHSTHEK-----PVG---IFPAFTSTNVSAALALDNRSTQSISLELKRAE 1568  
Db 1392 SITSKLSKVGTPRYAGKVKEKLEAGVNVKVDATTDKYNVARRLDPODDTGAVSPSKAE 1451

QY 1569 -----PVTNDISEL-----TSTLGKHFKDSATTKMLAALKELDDAKPAEQHLHQH 1616  
Db 1452 GKVTLPATPAGPKPGPLWDRGARTYGVGAKDSIT-----GPAGR---QGH 1494  
QY 1617 FSA-----KDVVGDERYEAVERNKLKLVIRQQA-----ADSHSMELGSASHSTYNN 1662  
Db 1495 LKVNADVNNNAVGEOSATAGKNGVALQVGQTQLTGGEIRSQOGKVELGSGVSQODVN 1554  
QY 1663 LSRINNOGVELLHKHFDAAALPASSAKRIGEMNNPALKDIKQLOSTPFSSASVSMEL 1722  
Db 1555 GQRYOGGGRVD-----AAATVGGLLG--AAKQSV--AGNVFPASGHASTQ 1597  
QY 1723 KDGLREQTEKAILDGK 1738  
Db 1598 AD-----AKAGVFSGK 1608  
RESULT 4  
TALI\_MOUSE  
ID TALI\_MOUSE STANDARD; PRT; 2541 AA.  
AC P26039;  
DT 01-MAY-1992 (Rel. 22, Created)  
DT 01-MAY-1992 (Rel. 22, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE TALIN.  
GN TLN.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OC NCBI\_TaxID:10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BALB/C; TISSUE-fibroblast;  
RX MEDLINE=9101390; PubMed=2120593;  
RA Rees D.J.G., Aides S.A., Singer S.J., Hynes R.O.;  
RT "Sequence and domain structure of talin";  
RL Nature 347:685-689(1990).  
CC -!- FUNCTION: PROBABLY INVOLVED IN CONNECTIONS OF MAJOR CYTOSKELETAL  
CC STRUCTURES TO THE PLASMA MEMBRANE. HIGH MOLECULAR WEIGHT  
CC CYTOSKELETAL PROTEIN CONCENTRATED AT REGIONS OF CELL-SUBSTRATUM  
CC CONTACT AND, IN LYMPHOCYTES, AT CELL-CELL CONTACTS.  
CC -!- SUBUNIT: BINDS WITH HIGH AFFINITY TO VINCULIN AND WITH LOW  
CC AFFINITY TO INTEGRINS.  
CC -!- PTM: PHOSPHORYLATED.  
CC -!- SIMILARITY: CONTAINS A DOMAIN FOUND IN BAND 4.1, EZRIN, MOESIN,  
CC RADIXIN, AND TALIN.  
CC -!- SIMILARITY: IN THE C-TERMINUS TO YEAST SLA2 AND C.ELEGANS ZK370.3.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: X56123; CAA39588.1; -  
DR PIR: S11661; S11661.  
DR HSP; P04002; IWFA.  
DR MGD; MGI:1099832; Tln.  
DR InterPro; IPR000299; -  
DR InterPro; IPR002558; -  
DR Pfam; PF00373; Band\_41; 1.  
DR Pfam; PF01608; ILWEPQ; 1.  
DR PROSITE; PS00660; BAND\_41\_1; 1.  
DR PROSITE; PS00661; BAND\_41\_2; 1.  
DR PROSITE; PS00662; BAND\_41\_3; 1.  
DR PROSITE; PS00663; BAND\_41\_4; 1.  
KW Structural protein; Cytoskeleton; Phosphorylation.  
FT DOMAIN 173 332 BAND 4.1-LIKE.  
FT VARIANT 1105 1105 L -> P.  
FT VARIANT 2180 2180 K -> M.  
SQ SEQUENCE 2541 AA; 269832 MW; 14EF75ABE9FC2CBB CRC64;





Db 2309 LIGAAAAIEAAAKLEQLPRAKPREADESLNFEQILEAAKSI-----AAATSALVKA 2362  
 QY 1725 GLREOTEKAILDGKVGREVGVLFDQRNRLRVKSVSVSVSKS-----EGFN 1772  
 Db 2363 ASAAQRE-LVAQCKYGAIPANAL--DDGWSOGLISAARMVAATNLCFAANAAYQVGA 2419  
 QY 1773 TPALLGTSNAAWGMERNIGTINFKYGQDQNTPRFTLEG-GIAQAN-----POVASA 1825  
 Db 2420 SOEKLISAKQVAATAQLLVACKYKADQDSEAMKRLQAGNAVKRASDNLVKAQAKAA 2479  
 QY 1826 LTDLKEGLEMK 1837  
 Db 2480 FEDQENETVVVK 2491

RESULT 5  
 YD86\_SCHPO STANDARD: PRT; 1957 AA.  
 AC Q10411;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE HYPOTHETICAL 222.8 KDA PROTEIN C1F3.06C IN CHROMOSOME I.  
 GN SPAC1F3.06C.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 OX NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-972;  
 RA Connor R., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;  
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: Z70690; CAA94624.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 1957 AA; 222785 MW; 3F480CA06171D9DA CRC64;

Query Match 2.3%; Score 217.5; DB 1; Length 1957;  
 Best Local Similarity 18.7%; Pred. No. 0.012;  
 Matches 386; Conservative 298; Mismatches 731; Indels 645; Gaps 93;  
 QY 33 SSSSPQNAASLAABGNKGRMPRIHOPST---AADGISAAHQQKSFSLRGCLGTKKFS 89  
 Db 18 SASSLVSSAASPFIIDSLETPRPNISRASTGQLAEDGDTSSQHEDSSEELK----- 68  
 QY 90 RSAPQGGPCTTHSKGATLRDLARDGGETOHEAAPDPAARLT-RSGGVRRNMDMAGRP 148  
 Db 69 ---RQEVGRMRHSDLSIDAKLGSSEGT-----ASSALPLTPRS-----PSNASWL 112  
 QY 149 MVKGGSGDKV-----PTQOKRHOLNFCOMQRTMLSKMAHPASANAGDRLOHSPPHIGS 204  
 Db 113 LVRGGLDSPILDINSVTQKSNLLNELKQVR-----SKLALEHENGILSLQ-----LSSS 163  
 QY 205 HHEIKEEVPVGTSKATTAAHADVIEAQEDDDSEFQO-----LHQORLAREPNNPPQPKLG 260  
 Db 164 NKKDK-----NTSSVTT-----LTSREDSVYFQKLTLMESNFSKQSEAYDLRSQLL 211  
 QY 261 VATPISAPQPKLTAVAESVLEGTDTTQSPKPKQSMKSGAGVTPPLATVLDKQKQLAP 320  
 Db 212 TVTEKLDKKEKDYEKIEDVSSIKASLAEQASNKSRLRGEQPRLEKLVSSNNK----- 264

QY 321 DNPPALNTLLKQTLG-----KTOHYLAHHASSDGSQSOHLLDNLKNGHLFDIKSTATSYS- 373  
 Db 265 -----TVST-LKOTENSLRAECKTLOEKLEKCAINEEDSKL-----LEELHNVANYS 312  
 QY 374 -VLHNSHPGEIKGLAQAGTSVSDGKSGIKSLSGSTQSHNKTMLSQPCEAHRSLITGI 432  
 Db 313 AIVH-----KDKL-----IEDLSTRISEFDNLKSERDT-LSIKNEKLEKLL--- 352  
 QY 433 WOHFAGAAARPOGES-IRLHDDKIHLHPGLGVQWQADKTHSOLSRQADGKLYAL-KDNR 490  
 Db 353 -RNTIGSLKUSRTNSQLEEEEMVEL-----KESNRTIHSLT-DAESKLSSFEQENK 402  
 QY 491 TLONLSDN-----KSSEKLVDKIKSVSDQGVAILTDTGRRHKMSIMPSLDASPESHI 545  
 Db 403 SLKGSIDYQNNLSKDKRWKOVSSQLEEARSSLAHAT---CK----- 442  
 QY 546 SLSLHFADAHQGLLHGKSELEAQSVIAISHGRVLVADSEGLFSAAIKPKQDGNELKWA- 604  
 Db 443 -----LAEINSEROFQNKIKDFEKIEODLRACLANSS-----NELKESA 483  
 QY 605 -----MPQHALDEHFGHDHIOISGFHDDHGOLNALVKN-NFRQOHACPLGNDHOFHPG 656  
 Db 484 LIDKKDQELNLRQIKQKQKVS---ESTQSSLSQSLQDILNEKKKHE----- 528  
 QY 657 WNLTDALVIDNOL---GLHHTNPEPHEILDMCHLGS-----LALQEGKLHYFDOLTGW 707  
 Db 529 -----VYESOLNELKGELOT-----EISNHEHSSQLSTLAAEKEAAVATNNEUSEK 576  
 QY 708 TGAESDCKQLKKGLDGAAYLLKDG-----VKRLMINOS-----TSSIKHGT 749  
 Db 577 NSLOTLCNAFOEKLAKSVQMLKENQFNSLDTSFKKLNESHQELNENHQTITKQLKDT 636  
 QY 750 ENVSFLPHVR-NKPEPGDALOGLNKDDKAQAMAVIGVKNKYLALTEKGDIRSFQ--TKPGT 806  
 Db 637 SKLOQLERANFEQKESTLSDENNDLRTKLLKLEESKSL-IKKOEDVDSLEKNTQTLK 695  
 QY 807 QQLERPAQTL-----SREGISGELKDIH----- 829  
 Db 696 EDLKSSEALRFSKLEAKNLREVIDNLKGKHETLQAEQNDLHSSLSDAKNTNAILSSELT 755  
 QY 830 ---VDHQKQ---NLVALTHEGEVFGHPREAWQNGAES--WHKL-----ALPQ 869  
 Db 756 KSEDEVKRLTANVETLTQDSKAMQSFSLVNSYQISNLYHELRRDHVNMSONNTLLE 815  
 QY 870 SESKLSDMSHEHKPIATFEDGSOHQLKAGWHAYAAPE----- 909  
 Db 816 SESKLT-DCENLTQNNMTLIDNVQKLM-----HKVNOESKVSSELKEYNGKLSLDLKNL 869  
 QY 910 RGPIAVGTSGSOTVFENLMO-----GVKGK-----IPSGGLTVKL 945  
 Db 870 RSSLNVAISDNDQILTQLAELSKNYDSLEQSAQLNSGLKSLAEAKQLLHTENEELHRL 929  
 QY 946 SAOTGGMTGAE-----GRKVSSEKFSERIRAYAFNPTMTSTPRPIKNAAYATOHGMOGREG 999  
 Db 930 DKLTGKLIKIESSDGLGKLTARQEE----- 956  
 QY 1000 LKPIYEMOGALIKOLDNAHVNRHNPAPQDLOSKL-ETLIDGEHGAELLNDMKRFRDELEOS 1058  
 Db 957 -----ISNLKEENMSQSAITSVKSKLDETL-----SKSSKLEADIEHLKKNVSEV 1002  
 QY 1059 ATRSVTVLGOHGV---LKSNGE-----INSEFKPSPGKA 1090  
 Db 1003 EVERNALLASNERLMDLLKNNGENTASLOTIEKKRAENDDLOSKLSVVSSEYE-----NL 1058  
 QY 1091 LVQSFNVRSGQDLSKSLQOAVHATPPSAESKLQSMGLGHFVSAGV---DMSHQKGEIPLG 1147  
 Db 1059 LLISSQTNKSLDEKTNQLKY-----IEKNVQKLLDEKQDNVELEELTSKYGL--- 1107  
 QY 1148 RQRPNDKTLTKSLRLDITVTIGELHELADKAKLVSDHKPADQIKQLRQOFDTLREKR 1207  
 Db 1108 -----GEENAOIKDELLALRKKSKQHDLG---ANFVDDLKEKSDALEQUTNEKNELIVSL 1160  
 QY 1208 YESNPVKHYTDMGFTTHNKALEANYDAVKAFINAFKK-----EHHGVNLTTRTTLVESQSGAE 1263

```
Db 1161 EQSN-----SNNEALVEERSDLANRLSDMKKSLSDSDNVISVIRSDLVNRVNDL 1210
Qy 1364 LAKLKNLTL-----LDS--GESMSFSRSYGGGVSTVPTLSK-KVPVPVI 1308
Db 1211 TLKDKDSLSTQYSEVQCDRLDLSDKGCESFNKY-----AVSLRELCTKSEIDVPSVI 1267
Qy 1309 -----PGAGITUDRAVNLFSRSTSGGLN-VSGGR---DGVSGVNMIVATGHDVMPYM 1356
Db 1368 LDDNFVFNAGFSELSRLTVLSLENVLDFAFNQVFKKMLD-----NRUTTTDAEFTKVV 1322
Qy 1357 TGKRTSAGNSDWLSAKHKISPLRIGAAVSGTLOQTQNSLK-FKLTEDLPFGFIHGLT 1415
Db 1323 ADLEKLOHEHDDWLQI-----GDLEKALDKSEKNFLRKEAEMTNIHSLE 1368
Qy 1416 HG---TLTPAELLQKIEHOMQKSGKLTFSVDTSANLDRAGINLNDGSKPNGVTVARVS 1472
Db 1369 EGKETKEKIEAELSSRLDQNLATKLNKQLD-----HLNQEIRLKEDEVILK-----EKES 1418
Qy 1473 AGLSASANLAAGSRERSTTSGQFGSTTSASNNRPTFLNGVG-AGANLTAALGVVAHSSTHE 1531
Db 1419 LIISLESL---SNQKRES-----SLDDAKNELEHMLDDTSSKNSSLMKIESINSSLDD 1471
Qy 1532 GKPVGIFPAFTSNVSAAL-----ALDNRTSQSLSLELKRAPVTSNDISBLTSLTGKHF 1587
Db 1472 -----KSFELASAVEKLGALQKLHSESL-----MENIKSOLQEAKEKIQV 1513
Qy 1588 DSATTMKLAALKELDDAKPAEQHLHQHFSAKDVGVVDERYEAVRNKLKLVIRQQAADSH 1647
Db 1514 DEST-----TOELDHETASK-----NNYEKLNKDKSDIIRDL-----1547
Qy 1648 SMELGSASHSTYNNLSRINDGIVELLKHFDAALPASAKRKLGMNNDPALKDIIKO 1707
Db 1548 -----ENIEQLNN-----LL-----AEKSAAVKRLSTE-----KE 1572
Qy 1708 LQSTPFSASVSMELKDLGRLQTEKAILDGKVGREEVGVLFQDRNRLRVKSVSVSQSVK 1767
Db 1573 SEILOFNRLADLEY-----HKSQVESELG-----RSKLKLASTTEELQALAE 1614
Qy 1768 SEGFTNPAALLGTSAAAMSERNITGTFNKYGQDQNTPRFTLEGGIAANPO-----1821
Db 1615 NERLSLTTMRDLQNOV-----KDLNLIKDSLEDLTLR--SLEDSVASLQKECKTKSN 1667
Qy 1822 ----VASALTDLKKEGLEMK 1837
Db 1668 TVESLQDVLTSVQARNAELE 1687

RESULT 6
AC ACAC_HUMAN STANDARD; PRT: 1781 AA.
AC Q02952; Q99970; O00498; O00310;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE A-KINASE ANCHOR PROTEIN 12 (A-KINASE ANCHOR PROTEIN 250 KDA) (AKAP
GN AKAP12 OR AKAP250.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
ON NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A. (ISOFORM 1).
RP TISSUE=Heart;
RC MEDLINE=97153077; PubMed=9000000;
RX Nauer J.B., Klauck T.M., Langeberg L.K., Scott J.D.;
RT "Gravin, an autoantigen recognized by serum from myasthenia gravis
RL patients, is a kinase scaffold protein.";
RL Curr. Biol. 7:52-62(1997).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Umbilical vein endothelial cells;
```

```
RX MEDLINE=98269042; PubMed=9604001;
RA Sato N., Kokame K., Shimokado K., Kato H., Miyata T.;
RT "Changes of gene expression by lysophosphatidylcholine in vascular
RT endothelial cells: 12 up-regulated distinct genes including 5 cell
RT growth-related, 3 thrombosis-related, and 4 others.";
RL J. Biochem. 123:1119-1126(1998).
RN [3]
RP SEQUENCE OF 43-1781 FROM N.A.
RC TISSUE=Umbilical vein endothelial cells;
RA Bowditch R.D., Ginsberg M.H.;
RT "Sequence of gravin cDNA isolated from a human umbilical vein
RT endothelial cell library.";
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 1477-1781 FROM N.A.
RC TISSUE=Umbilical vein endothelial cells;
RX MEDLINE=92395179; PubMed=152245;
RA Gordon T., Grove B., Loftus J.C., O'Toole T., McMillan R.;
RT "Molecular cloning and preliminary characterization of a novel
RT cytoplasmic antigen recognized by myasthenia gravis sera.";
RL J. Clin. Invest. 90:992-999(1992).
CC -!- FUNCTION: ANCHORING PROTEIN THAT MEDIATES THE SUBCELLULAR
CC COMPARTMENTATION OF PROTEIN KINASE (PKA) AND PROTEIN KINASE C
CC (PKC).
CC -!- SUBUNIT: BINDS TO DIMERIC RII-ALPHA REGULATORY SUBUNIT OF PKA.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC. MAY BE PART OF THE CORTICAL
CC CYTOSKELETON.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS: 1 (SHOWN HERE) AND 2; MAY BE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN ENDOTHELIAL CELLS, CULTURED
CC FIBROBLASTS AND OSTEOSARCOMA, BUT NOT IN PLATELETS, LEUKOCYTES,
CC MONOCYTIC CELL LINES OR PERIPHERAL BLOOD CELLS.
CC -!- INDUCTION: ACTIVATED BY LYSPHOSPHATIDYLCHOLINE (LYSOPC).
CC -!- DOMAIN: POLYBASIC REGIONS LOCATED BETWEEN RESIDUES 265 AND 556 ARE
CC INVOLVED IN BINDING PKC.
CC -!- DISEASE: ANTIBODIES TO THE C-TERMINAL OF GRAVIN CAN BE PRODUCED BY
CC PATIENTS WITH MYASTHENIA GRAVIS (MG).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; U81607; AAC51366.1; -
DR EMBL; AF001504; AAB58938.1; -
DR EMBL; AB003476; BAA19927.1; -
DR EMBL; M96322; AAA35931.1; -
DR MIM; 604698; -
DR PIR; A43922; A43922.
DR InterPro; IPR001573; -
DR InterPro; IPR001899; -
KW Antigen; Alternative splicing.
FT DOMAIN 603 633 AKAP 1.
FT DOMAIN 752 782 AKAP 2.
FT DOMAIN 797 827 AKAP 3.
FT DOMAIN 98 101 POLY-GLU.
FT DOMAIN 265 556 INVOLVED IN PKC-BINDING (PROBABLE).
FT DOMAIN 1540 1553 RII-BINDING (PROBABLE).
FT VARSPLIC 1 98 MISSING (IN ISOFORM 2).
FT VARSPLIC 99 106 EEEVIVTE -> MLGTITIT (IN ISOFORM 2).
FT CONFLICT 117 117 E -> K (IN REF. 2).
FT CONFLICT 142 144 NRN -> TPEI (IN REF. 2 AND 3).
FT CONFLICT 215 215 Q -> K (IN REF. 2).
FT CONFLICT 448 448 G -> E (IN REF. 2 AND 3).
FT CONFLICT 694 694 R -> G (IN REF. 2 AND 3).
FT CONFLICT 867 867 G -> S (IN REF. 2 AND 3).
FT CONFLICT 986 986 S -> A (IN REF. 2 AND 3).
FT CONFLICT 1530 1530 E -> EE (IN REF. 3 AND 4).
FT CONFLICT 1581 1581 V -> M (IN REF. 4).
```

FT	CONFLICT	1501	1601	Q -> L (IN REF. 2).
5Q	SEQUENCE	1781	AA; 191439 MW; BA813937379FAC0F CRC64;	
	Query Match	2.38;	Score 214; DB 1; Length 1781;	
	Best Local Similarity	17.98;	Pred. No. 0.015;	
	Matches 339; Conservative	278;	Mismatches 718; Indels 564; Gaps	
Qy	29	QGGSSSSPQNAASLAAEGKNRKMPIRIIHQPSTAADG-----ISAHQOKKSFSLRGCL 83		
Db	245	BQSHAEISPPAESQGAVEECKEEGKQEKPEKSAESPTSPVTSETGSTFKFFQTQGA 304		
Qy	84	GTKK--FIRSAPOGQGTTHSKGATLRDLLARDGDTQHEAAAPADARLTRSGGVKRN 140		
Db	305	GWRKTSFRPKDEVEASKKKQDEPKVDTEDEGK-----AEVASEKLTASEQAHPQE 359		
Qy	141	MDDMAGRPVMYKGGGDEKVPVTOQKRHOLNNFGOMQROTMLSKMAHPASANAGDLRQHPH 200		
Db	360	PAESAHEPRLSAEYKVELPSEEQ-----383		
Qy	201	IPGSHHEIKPEPVGTSKATTAAHADRVEIAAQEDDDSEF-----QOLHOORLARERE--N 252		
Db	384	VSGQGPSEPKP---APRLATEVDEKTEVHQEEVAAEVHVSTVEERTTEOKTEVEETAGS 440		
Qy	253	PPQPKLGV-ATPISARFQKLTAAVGSVLEGGDTTQ-SPLKP-OSMLKSGAGV--TP 306		
Db	441	VPAEELVGMDAEPOEAEPANELVKLKTVCVSGEDPTGGADLSPEKVLSPPEGVVSEVE 500		
Qy	307	LAVTLDRGKGLQALPDNPALNTLLKQPLGK-----DTQYLAHHASSDGSQHL 355		
Db	501	MLSSQERMKYQGSPLKKLFTSTGLKLSGKQKQKRGKGDEESGEHTVPADSPDQSE--558		
Qy	356	LDNKGHLFDIKSTATSYSLVHNSHPGKIKGLAAGTGSVSDGKSKTSLSGTOSHNK 415		
Db	559	-EQKG-----ESSASS-----PEEPEI--TCLEKGLAEVQDGEAE--GATSGEKK 602		
Qy	416	TMLSQCEAHRSLTLGIWHPAGAARPQGESIRLHDDKTHILHPELGVWQADKDTHSOL 475		
Db	603	REGVTPWASFKKMVT-----PKKRVRRPSES-----DKELDKVKSATLSSTESTASEM 652		
Qy	476	SROADGKLYALKNRNTLQNLSDNKSSEKLY-----DKIKSYSYVDORQGVAI-----L 522		
Db	653	QEBMGKSVSEPKPEEPKRKYDTSVSWEALLICVSSKKRRRRSSDEEGPRAMGGDHOK 712		
Qy	523	TDTPGRIHKMSIMPSLDASPSHISLSLHFADAHQG-----LHKGSELEAQSVATSHGR 577		
Db	713	ADEAGKDKETGTDGILAGSQEH-----DPQGSSSPQEQAGSPTEGEGVSTWESFKRL 764		
Qy	578	VVADSEGLKLSAALPKOGDGNELKMKAMPQHALDEHFGHDHQITSGFFHDDHQLNALVKN 637		
Db	765	VT-----PRKKSXKLEEKs-----EDSTAG-----785		
Qy	638	NFRQOHACPLGNDHQHFPGWNLTDALVIDNLGLHHTNPPEHILDMGHGLSLAQEGLK 697		
Db	786	-----SCVEHSTPDTE-----796		
Qy	698	HYFDQLTKGWTGAESDCKQLKGLGGAAYLLKDGVEKVRLLNIQSTSSIKHGNTENVSUPH 757		
Db	797	-----PGKEESWVSIKKIPGRKKRRKPDGKQEQAPVEDAGPT-----833		
Qy	758	VRNKPEPGDALQGLNKDKDAQAAAVTGVNKKYLALTEKGDIRSFQTKPGTQOLERPQATLS 817		
Db	834	-----GANEDD-SDPVAVPLSEYDAY-EREKMEAQQAQKGAEQPEQKAAT--877		
Qy	818	REGISGELKDHYVDHKONTIYA-LTHEGEVFHQPREAWGAESSWHKL--ALPQSESK 873		
Db	878	--EVSKELSQSQVHMAAAVADGTRAATITIEERSPSWISASVTEPLEQVEAEAAALLTEEV 935		
Qy	874	LKSILDSHHEKPIATFEDGSOHLKAGWHYAAAPERGPLAYCTSGSTVFNRLMQGVK 933		
Db	936	LREYVTAEEPPIVT-----BPLPE-----NREARG--D 962		
Qy	934	KVIPGSGLTAVK--LSAQTGWTGA-EGRKVSXKFSERIRAYAFNPTMTPRIKNAAVT 990		

OS Bordetella pertussis.  
 OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;  
 OC Bordetella.  
 OX NCBI\_TaxID=520;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RX MEDLINE=90355839; PubMed=2388559;  
 RA Reiman D.A., Domenighini M., Tuomanen E., Rappuoli R., Falkow S.;  
 RT "Genetic characterization of Bordetella pertussis filamentous  
 haemagglutinin: a protein processed from an unusually large  
 precursor.";  
 RL Mol. Microbiol. 4:787-800(1990).  
 RN [2]  
 RP SEQUENCE OF 1-3261 FROM N.A.  
 RX MEDLINE=89202384; PubMed=2539596;  
 RA Reiman D.A., Domenighini M., Tuomanen E., Rappuoli R., Falkow S.;  
 RT "Filamentous haemagglutinin of Bordetella pertussis: nucleotide  
 sequence and crucial role in adherence.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 86:2637-2641(1989).  
 CC FUNCTION: EVIDENCE FOR A ROLE IN HOST-CELL BINDING AND  
 INFECTION.  
 CC -!- SUBCELLULAR LOCATION: SURFACE.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; M60351; AAA22974.1;  
 DR EMBL; M60351; AAA22975.1; ALT\_INIT.  
 DR EMBL; M60351; AAA22976.1; ALT\_INIT.  
 KW Antigen; Hemagglutinin.  
 KW SEQUENCE 3591 AA; 367420 MW; BF7418B30D6E5138 CRC64;

Query Match 2.2%; Score 207.5; DB 1; Length 3591;  
 Best Local Similarity 19.3%; Pred. No. 0.09;  
 Matches 427; Conservative 268; Mismatches 792; Indels 725; Gaps 114;  
 QY 5 SLGTEHKAHVTAHNPNVGHVALQOGSSSSPQNAASLAEGKNGKMPRIHQPTAA 64  
 DB 1379 SLDIKKGAAQVTVAGRYAEHGEVSIQDYTSAD--AIALAAQVTRG-----GA 1426  
 QY 65 DGISAHQKKSFSRLGCLGCTKFSRSPQGPQCTTH-SKGATLRDLARDGETQHEAA 123  
 DB 1427 ANLTSRHDTFRFNKIR-LMGPLOVNAAGPVSNTGNLKVREGVTV--TAASFDNETGAEM 1483  
 QY 124 APDAARLTRSGGKVRNMDMAGRPMVK-----GSGG 155  
 DB 1484 A-KSATLTTSGAARN-----AGKMQVKEAATIVAAVSNSPGTFTAGKDIIVTSRGGFDN 1536  
 QY 156 EDKVTQO---KRHQLNN-----FGOMRQTMLSKMAHPASANAGDRLOHS 197  
 DB 1537 ECKMESNKDIVTKTEQFSNGRVLDAKHDLTVTASGQADNRSGLKAGHDFTVQA--ORDINS 1595  
 QY 198 PHPIGSHHEIKEEPVGSYKATTAHADRVIEIAQEDDSEFQQLHQORLARENPPOPP 257  
 DB 1596 GPMAGHDATLAKAPHLNTEQVVAGH-----DIHINSKLENTGRVDA 1639  
 QY 258 KIGVATPISARFOPKLTAVAE-----SVLEGTD-----TTOSPLKP--OSMLKSGAGVTP 306  
 DB 1640 RNDIALDV-ADETNTGSLVAEHDATLTLAQGTORDLVVDQDHLIPVAEGTLRVKAKSLTT 1698  
 QY 307 LAVTLDKGL---OLAPNPAPALNTLLKOTLKGKDTCHYLAAHASSDGSOHLILDKNKGHL 362  
 DB 1699 ELETGNPGSLIAEVQENINQKAI-----VVGKDLTSSA-HGNVANEANALLWAAGEL 1751  
 QY 363 FDIKSTATSYVHLNHPGEIKGKLAQAG-----TGSVSVVDGKSGKISLG-----SGTQS 412  
 DB 1752 -----TVKQAQNTNK-----RAALTEAGGNARLTAAVALLNKLGRIRAGEDMHLDPRI 1800

QY 413 HNKTMLSPQGEAHS-----LLTGIWHPAGAAAR----- 442  
 DB 1801 ENTAKLS--GEVORKGVQDVGGEGHGRWSGIGYVNYWLRAGNGKAGTIAAPWYGDLTA 1858  
 QY 443 -----QGESIRLH-----DDKTHILHPGLVQWQSKADKTHSLSRQADGK----- 482  
 DB 1859 EOSLIEVGRKDLVYNAGARKDEHRLHNE--GVIOAGG---HGHIGGDVNRKSVVTVTSAM 1913  
 QY 483 -----LVALKDNRTLQNLSDN-KSSEKLVDKIKSVSDQDQGVAILTDTDPGRHKM 531  
 DB 1914 EYFKPLPVSLTALDNRAGLSPATWFOSTYELD-----YLLDNRYEYINGLYPTYTEM 1969  
 QY 532 SI-----MPSLDASPESHISLSHFADAHQ-----LLHG---KS 563  
 DB 1970 SVNTLKNLDLGYQAKPAPTAPPMPKAPD--LDLRGHTLESAGRRKIFGEYKLOGEYKA 2027  
 QY 564 ELEAQSV-AISHGRVLVADSEGLFSAAI-----PKGDCGNELKMKAMPOHALDEHFGH 616  
 DB 2028 KMAVQAVEAYGEATRRVHDQIGORYKALGMDMETREVDG-----IIOEFAA 2075  
 QY 617 DHOISGFFHDDHGQLAA---LVKNFRQO--HACPLGNDHOFHPGWNLTDAVIDNQLG-- 670  
 DB 2076 DLRTVYAKQADQATIDAETDKVARYKSQIDAVRL---QAIQPG-RVTLAKALSALGAD 2131  
 QY 671 ---LHHTN-----PEPHEILDMCHGLSGLALOEKGLHVFOLDTKGW 707  
 DB 2132 WRALGHSQLMORWKDFKAGRGAEIAFYPKBQTVLAAG--AGLTLSNGAIH-----NGE 2183  
 QY 708 TGAESDCKO--LKKG-----LOGAAYLLKD-GEVKRLININOSTSSIKHGTENVSFLPHV 758  
 DB 2184 NAAQNRPRPEGLKIGAHSAITSVSGFDALRDVLEKLRIDIDDAALAA-----VLNPHI 2236  
 QY 759 -----RNKPE-----PGDALQGLMKDDKAQAMAYI 783  
 DB 2237 FTRIGAAQTSADGAAGPALARQAPQAPETGMVMDARGLSADALASLASLDAAGLEYS 2296  
 QY 784 GVNKYIALTEKDIRSQIKPGTOQLERP---AOTLSREGISGLKELK---IIVDHKQNLV 837  
 DB 2297 G-RRNAQVADAG-----LAGPSAVALPAGVADGVPEVPTGDQDQVPAVGLQBPVA 2348  
 QY 838 ALTHEGEVPHQPREAMQNGAESSSMHKLALPOSSEKLSLDMSHHFKPIATPEDGSOHOL 897  
 DB 2349 TVRVAPPAVALPR-----PLFETRIKFIQDSKFY-----GSRYFF 2383  
 QY 898 KAGGHAYAAAPERGLAVTSGSTQVF-----NRLMQGVKGV-IPSGSLTVKLSAQTTG 951  
 DB 2384 EQIGY---KPDRAARVAGDNYFDTTLVREQVRALGGYERLPRVGVVALVAKLMDSAGT 2439  
 QY 952 MTGAERKV-----SSKFSERIRAYAFNPTMSTPRPIKNAAYATOHGMOGREGKLPY 1004  
 DB 2440 VGRALGLKGVAPTAQQLKQADRFVWVTVIDGQKVLAPRLYTE---ATROGQITDQY 2496  
 QY 1005 EMQGLALIK-----QLDAHNVRHNAFPQDLQSKLETLDLGEHGAELNDMKRFRDEBQ 1057  
 DB 2497 AGGALLIASGDDVTNTDGHV---SSVNGLIQGRSVKVDAGK-GKVWVADSKGAGGIE- 2552  
 QY 1058 SATRSVTVLQHGVLKSNGEINSEFPKSPKALVOSFNVRNSGQDLSLQOAVHATPP 1117  
 DB 2553 -ADEVDVSGRDIGI--EGGKLR-----GKDVRLKADTVKVIATSM 2589  
 QY 1118 SAESKLQSLMGLHGFVSAGVD-MSHOKGEIPLGRQRPNDKALT---KSRILDTVTITGEH 1174  
 DB 2590 RYDDK-----GRLAARGDGLDAGGGUHIIEAKRLETAGATLKGKVLDDVDVGLGGVY 2644  
 QY 1175 ELADKAKLVSDHPDQAIKQLROQFDTLREKRYESNPVKHYTDMGFTH---NKALEANY 1231  
 DB 2645 EAGSS-----YEN-----KSSTPLGSLFAILLSSTTETNQ 2673  
 QY 1232 DAVKAFINAFKKBHHGVNLTTRTV-----LESQGSALAKKLKNTLLSLDSESMSFSR 1285  
 DB 2674 SA-----HANHYGTRIEAGTLEGKMNLEIGGSVDA---AHTDLSVARDARFKAAA 2722

```
Qy 1286 SVGGGVSTVFVPTLSKKVPPVPIPGAGITLDRAYNLSFRS--GGLNVSFGDRGGVSGNI 1344
Db 2773 DFAHEHEKDVQLSLGAKV-----GAG-----GYEAGFSLGSSGLEAHAGR--GMTAGA 2771
Qy 1345 MYATCHDVMYPT-----GKKTSGAGNASDWLSAKHKISPDL---RIGAAV 1386
Db 2772 EVKGVYRASHEQSSETEKSYRANLNFGGSGVEAGNVLDIGGA-----DINERNYGGAA 2825
Qy 1387 SCTLOGTIONSLLKFKLTDELPGFTHGLT-----HGTLPALLOKQGI 1429
Db 2826 KGN-AGT-EEALURAKKVKSTKYVSEQTSQSGMSVEVASTASRSLLTATIRLGDSV 2883
Qy 1430 EHOMKQSGKL-----TFSDVTSANLDRAGINLNDGSKPVGTA----- 1469
Db 2884 AQNVDEGREIRGELMAAQVAEATQVLTADTAA-VALSAGISADFDSSHSRSTISQNTQYL 2942
Qy 1470 -----RVSAGLSASANAAGSR---ERSTTSGQFGSTTSASN 1503
Db 2943 GGNLSIEATEGATLVGAKFGGDDQVSLKAASKSVNLMMAAESTFESYSESHNPHASADANL 3002
Qy 1504 NRPTFLNGVGAGANLTAALGVAHSSTHE-CKPVGIFPFTSTNVSA--LALDNRSTQSIS 1561
Db 3003 GANAVOGAVGLG--LTAGMGTSHQITNETGK-----TYAGTSVDAANVSID--AGKDLN 3052
Qy 1562 LELKRAEPTVNSDISETLTSTLGHKFKDSATTKMLAALKKELDDAKPAEQHLILOOHFSKAD 1621
Db 3053 LSGSRV-----GKHVVLDVEGDINATSKQ----- 3077
Qy 1622 VVGDER-YE-----AVRNKLKLVIRQQAADSHMELGSA--SHSTYNNLSRI 1666
Db 3078 ---DERNTNSSGGGWDASAGVAIQN-RTLIV-----APVGSAGFNFTEDHN-SRL 3122
Qy 1667 NNDGLVELL-----HKHFDALPASSAKRLGEMMNNDPALKDIKQLQSTPFSSASVS 1719
Db 3123 TNDGAGVVASDGLTGHVKGANDANLTGATTDLGSGKN---LK-----VDCAVNA 3168
Qy 1720 MELKDLREQTEKAILDGVKREEVGVLFQDRNNLRKVSQSVSKSGEGFNTPALLLG 1779
Db 3169 ONLKD-YRDK-----DGGSGGLNVG-----ISSTTLAPTGVAVFG-----RVAG 3206
Qy 1780 TSNSAAMSERNIGTINFKYGQDQNTPRFTLEGGTAQANPQVASALTDLKK 1831
Db 3207 EDYQAEQRATIDVG-----QTKDPAFLQVGGGVKGTLNQDAQAATVVOR 3250

RESULT 8
APC_MOUSE
ID APC_MOUSE STANDARD: PRT: 2845 AA.
AC Q61315: Q62044;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ADENOMATOUS POLYPOSIS COLI PROTEIN (APC PROTEIN) (MAPC).
GN APC.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1 AND 2). AND VARIANTS.
RC STRAIN=C57BL/6J, AND CAST/EI; TISSUE=Brain;
RX MEDLINE=92263101; PubMed=1350108;
RA Su L.-K., Kinzler K.W., Vogelstein B., Preisinger A.C., Moser A.R.,
RA Luongo C., Gould K.A., Dove W.F.;
RT "Multiple intestinal neoplasia caused by a mutation in the murine
RT homolog of the APC gene.";
RL Science 256:668-670(1992).
RN [2]
RP ERRATUM.
RA Su L.-K., Kinzler K.W., Vogelstein B., Preisinger A.C., Moser A.R.,
RA Luongo C., Gould K.A., Dove W.F.;
RA Science 256:1114-1114(1992).
RN [3]
```

```
RP SEQUENCE OF 1-45 FROM N.A.
RC STRAIN=BALB/C; TISSUE=Liver;
RA Dicker F., Lambert S., Reitmaier A., Ballhausen W.G.;
RT "The murine APC gene: alternative splicing of 5' untranslated
RT region segments.";
RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
[4]
RP ALTERNATIVE SPLICING.
RX MEDLINE=94061824; PubMed=8242607;
RA Oshima M., Sugiyama H., Kitagawa K., Taketo M.;
RT "APC gene messenger RNA: novel isoforms that lack exon 7.";
RL Cancer Res. 53:5589-5591(1993).
CC -I- FUNCTION: TUMOR SUPPRESSOR. ALLOWS THE RAPID TURNOVER OF BETA-
CC CATENIN. APC ACTIVITY CORRELATED WITH ITS PHOSPHORYLATION STATE
CC ALLOWS THE DOWNREGULATION OF CYTOPLASMIC BETA-CATENIN (BY
CC SIMILARITY).
CC -I- SUBUNIT: FORMS HOMODIGOMERS AND ASSOCIATES WITH CATENINS (BY
CC SIMILARITY).
CC -I- ALTERNATIVE PRODUCTS: 4 ISOFORMS; 1 (SHOWN HERE), 2, 3 AND 4; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -I- TISSUE SPECIFICITY: EXPRESSED IN LIVER, SPLEEN, KIDNEY, HEART,
CC LUNG, BRAIN, STOMACH, INTESTINE, TESTIS AND OVARY.
CC -I- PTM: PHOSPHORYLATED BY GSK-3B (BY SIMILARITY).
CC -I- SIMILARITY: CONTAINS 7 ARM REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch)
CC -----
CC EMBL: M88127; AAB59632.1; -
CC EMBL: U02937; AAA03443.1; -
CC HSP: Q02248; 2BCT.
CC MGD: MGI:88039; Apc.
CC InterPro: IPR000225;
CC Pfam: PF00514; Armadillo_seq; 4.
CC PROSITE: PS00176; ARM_REPEAT; 1.
CC Anti-oncogene. Phosphorylation; Alternative splicing; Repeat;
CC Coiled coil.
KW FT DOMAIN 1 61 COILED COIL (POTENTIAL).
FT DOMAIN 125 245 COILED COIL (POTENTIAL).
FT DOMAIN 451 493 LEU-RICH.
FT REPEAT 503 545 ARM 1.
FT REPEAT 546 589 ARM 2.
FT REPEAT 590 636 ARM 3.
FT REPEAT 637 681 ARM 4.
FT REPEAT 682 723 ARM 5.
FT REPEAT 724 765 ARM 6.
FT DOMAIN 739 2834 ARM 7.
FT DOMAIN 1130 1156 SER-RICH.
FT DOMAIN 1556 1575 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 1864 1891 ASP/GLU-RICH (ACIDIC).
FT VARSPLIC 243 276 HIGHLY CHARGED.
FT VARSPLIC 310 410 MISSING (IN ISOFORM 2 AND ISOFORM 4).
FT VARIANT 120 120 MISSING (IN ISOFORM 3 AND ISOFORM 4).
FT VARIANT 493 493 T -> A (IN STRAIN CAST/EI).
FT VARIANT 493 493 V -> I (IN STRAIN CAST/EI).
FT VARIANT 797 797 Y -> F (IN STRAIN CAST/EI).
FT VARIANT 1330 1330 A -> T (IN STRAIN CAST/EI).
FT VARIANT 1618 1618 A -> S (IN STRAIN CAST/EI).
FT VARIANT 2294 2294 G -> A (IN STRAIN CAST/EI).
FT VARIANT 2496 2496 H -> Q (IN STRAIN CAST/EI).
FT VARIANT 2523 2523 T -> A (IN STRAIN CAST/EI).
FT VARIANT 2813 2813 T -> S (IN STRAIN CAST/EI).
SQ SEQUENCE 2845 AA; 311086 MW; 145CA73CF570A499 CRC64;
```

Query Match

2.2%; Score 204; DB 1; Length 2845;

Best Local Similarity 19.9%; Pred. No. 0.093;

Matches 412; Conservative 272; Mismatches 808; Indels 580; Gaps 97;











```
Qy 1153 ND-----KTALTKSLIILDTVTIGELHELADKAKLVSDHKPDADQIKQLROOFDTLREKR 1207
Db 1166 NTVYDAANSTFSAVAVNADYAAQVDLYADKAH-----TQLKHFDT---KV 1210
Qy 1208 YESNPVKHYTDMGFTH--NKALEANYDA-----VKAFINAFKKEHH----- 1246
Db 1211 RLTAFT--FTDLKFNNGSDQISEATIKVTGTVSSDTKTVNVGDTVAALDAQHFFSVDPV 1268
Qy 1247 --GVNLTRTVLESOGSALAKK-----LKNLLSLDSGSMFSRSYSGGVSTV 1294
Db 1269 NYGDMTIKVTATDEGNTTTEQKTTTSYDPPDLKNA-VTFDQG--VKFCANEFNATSAK 1325
Qy 1295 F-----VPTLSKKVPVPV---IPGAGITLDRAYNLFSRTSGGLNVS-FGRDGGVSGN 1343
Db 1326 FYDPKGTGATIGVKVHPHTTTLQVQKQIKNDITFSEFLDLGLGQKPPG----- 1377
Qy 1344 IMVATGHDVMPYMTGKTSAGNASDWLSAKHKISPDRLICAAVSGTLOCTLQNSLKFKL 1403
Db 1378 --VVVGD-----TTQNKTFQEALTFLDA---VAPTLSD-----SSTDAPVYNDPNFQIT 1424
Qy 1404 EDELPFGTHGLTLP-AELLQGI--EHOMKQSKLTFVSDTSANLDRAGINLNE- 1459
Db 1425 -----GTADNAQVLSLAINGSHVASQYADININSGKPGHMAIDQPVKLLG 1471
Qy 1460 -----DGSKPNCVTVARVSAGLSANLAAGSRERSTT-----SQFGSTTS 1500
Db 1472 KNLVTAVTDSNNTTTKITVYEPKTLAAPTVPSTTEPAKTVTLTANAAATGETVQ 1531
Qy 1501 -ASNNRPTFLNGVAGANLTAALGVASHSTHEGKPVGIF---PA--FTSTNVSAALALD 1553
Db 1532 YSADGKTYQDPAAGVTVA-----NCTPFKSTDLYCNESPAVDYVYVNLKA----- 1580
Qy 1554 NRTSISILEKRAPEVTSNDI--SELTSTLKGHKDSATTMLAALKELDDAKPAEQULH 1611
Db 1581 -----DDPAQLQTAQALTNLIASAKTLSASGKY--DDATTTTALAA-----ATQKAQTA 1627
Qy 1612 ILQHFSAKDVVDERYEAVRNKLKLVIRQQAADSHSMELGSASHSTTYNNLSRINDGI 1671
Db 1628 LDQTDASVDSLUG-----ANRLQ-----TAINOLA----- 1653
Qy 1672 VELLKHKFDAAALPASSAKRLGEMMNDPALKDIIKQLQSTPFSSASVSMELKDLREQTE 1731
Db 1684 -----AKLPADKKTSL-----LNQLQSV---KAALGTDLGNQDTPSTG 1688
Qy 1732 K---AILDGKVGREVGVLFDQRNRLRVKSVSVSQSVKSEGFN--TPALLIGTSNAA 1785
Db 1689 KTFTAALDDLVAQAQAGTQTADQLQASLAKVLDAVLAKLAEGTKAATPA-EVGNAKDAA 1746

RESULT 11
KEND_HUMAN
ID KEND_HUMAN STANDARD: PRT: 3321 AA.
AC O95613; 043152;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE KENDRIN.
GN KIAA0402.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi.
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Li Q., Joshi H.C.;
RT "Kendrin, an integral component of centrosome.";
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1512-3321 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98116655; PubMed=9455477;
RA Ishikawa K.-I., Nagase T., Nakajima D., Seki N., Ohira M.,
```

```
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. VIII.
78 new cDNA clones from brain which code for large proteins in
vitro.";
RL DNA Res. 4:307-313(1997).
CC -|- FUNCTION: CENTROSOMAL PROTEIN.
CC -|- SIMILARITY: STRONG, TO MOUSE PERICENTIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
DR EMBL: U52962; AAD10838.1; -.
DR EMBL: AB007862; BAA23698.1; -.
DR HSSP: P02649; ILPE.
KW Coiled coil.
FT DOMAIN 258 553 COILED COIL (POTENTIAL).
FT DOMAIN 675 835 COILED COIL (POTENTIAL).
FT DOMAIN 999 1135 COILED COIL (POTENTIAL).
FT DOMAIN 1288 1938 COILED COIL (POTENTIAL).
FT DOMAIN 2053 2071 COILED COIL (POTENTIAL).
FT DOMAIN 2525 3075 COILED COIL (POTENTIAL).
FT CONFLICT 1523 1523 H -> Q (IN REF. 2).
FT CONFLICT 2177 2177 R -> M (IN REF. 2).
FT CONFLICT 2538 2538 T -> A (IN REF. 2).
FT CONFLICT 2828 2906 MISSING (IN REF. 2).
FT CONFLICT 3012 3012 K -> E (IN REF. 2).
FT CONFLICT 3125 3125 S -> Y (IN REF. 2).
FT CONFLICT 3289 3289 G -> E (IN REF. 2).
FT CONFLICT 3309 3321 YYQLLQKNPATR -> VLPDSTSKSKSCHPMIKQ (IN
REF. 2).
FT SEQUENCE 3321 AA; 376327 MW; 0DB3D287BF7D8762 CRC64;

Query Match 2.1%; Score 195; DB 1; Length 3321;
Best Local Similarity 17.0%; Pred. No. 0.32;
Matches 290; Conservative 256; Mismatches 588; Indels 570; Gaps 72;

Qy 377 NSHPGEITKGKLAQAGTGSVSDGSGKISLGSCTOSHNTMLSQPGEAHRSLLITGWQHP 436
Db 110 NDHPE-----QCGMFTVS-----DHPEQHGMTVG--DHP 139
Qy 437 AGAARPOGESIRLHDDKTHILHP--ELGVQMSADKDTSHLSQSRQADGKLYALKNDRTLN 494
Db 140 -----PEORGMTVSD-----HPPEQHGMTVSDHP-----PEQRGMFT 173
Qy 495 LSDKNSSEKLYDKIKSVSDQRGQVAIILDTGPRHKMSTMPSLDASPESHIS----- 546
Db 174 ISDHQPEORGMTVSDHTPEQRG--IFTISDHPAEOGDMFTKECEQECELAITDLESGRD 232
Qy 547 -LSLHFADAHQGLHKGSELEAQSVASHGRVLVADSEKGLFSAALPKQDGNELMKAM 605
Db 233 EAGLHQSQAQVHGL-----ELEALRLSLN-----MHTAQLELTQANLQKEKETALTRE 283
Qy 606 PQHALDEHFHDHQLISGFPHDDHQLNALVKNFNQOQHACPLGNHQHFGHPGNLTALVI 665
Db 284 -----LNSRRAQELALLQS--RQOHELE----- 305
Qy 666 DNQLGLHHTNPPEHILDMGHLS-----LALQEGKLHYFDQLTKGWTGAESDC-KOLK 718
Db 306 ---LREQHAREKEEVVLCGQGAELKEKLOSEMKNQAIIVTLKEDWSEKDLLENLR 362
Qy 719 KGLDGAAYLLKDGVEVKRLNINQSTSSIKHGT-----NVFSLPHVKNKPEPGDALGLNK 773
Db 363 KEL-----SAKHOSEMEDLQNFQKELAQRAELEKIFODKNQ 400
Qy 774 DDKAQAAMAVIGNKYALATEK--GDIRS-----FOIKPGTQOLBERPAOTLSREGI 831
Db 401 AERALRNL---ESHQAAAIKLRDLQSHGRCLDELFKFKSEKE-----KOLELENL 452
```

QY	822	SGELKDIHYDHHKONLYALTHEGEVFIHQPREAWONGAESSSWHKLALPQSSKIKSLDMSH	881
Db	453	QASYEDLKAQSOEETIRNL-----WSOLPSARTSROFELSHEHQLLART--SRVEDLSQLK	505
QY	882	EHKPIATFEDGSOHOLKAGGHAYAAPEKPLAVGTSGSTVENRLMOGVKGKVPICPSSL	941
Db	506	Q-----REKTOHSELEQLRIYFEKKLRDAEKTQOEDUTLLQORLOGGAREDALDS-V	557
QY	942	TVKLSAQGTGMTGABRK-----VSKTFSERIRAYAFN-----PTMSTPRPIKNA	986
Db	558	EVLGSCVGLKEKPEKGRKHVDLEPERHKESLPREQAELEESHRHQLEALESPCLQIHE	617
QY	987	AY-----ATQHW-----QGREGLKPLYEMOGALIKQIDA-----H	1017
Db	618	GHVSDRCCVETGALGHEWLEPSEGHSQLPWHVHLOGVQDGLDEADTERAARVLGLETEH	677
QY	1018	NVRHNAPODLSOKLETLB-----LGEHGAELN-----	1046
Db	678	KVOLSILQTELKEETELLKIENRNLFEKLOHETRLKDDLEKVKHNLIEDHQELNNAKQK	737
QY	1047	-----DMKRFRDELEQSASTRVTVL-----QOHOGVLKSGEINSEFK	1084
Db	738	TELMKQEFQKQETDMVKMEQLQREAEELITLMLELREKAESKQOTIINKLELREAEMR	797
QY	1085	PSPGKALVQSFVNRSGDLSKSLQ-----AVHAT-----PPSAES-----	1121
Db	798	QLQDQQAQIILDELSLTQQARLOQLQEOULTSDDALHCSQCGRPEPTAODGELAALHYK	857
QY	1122	---KLQSMI--GHFVSAGVDM-----HOK-----GEIPLGRQRDPN	1153
Db	858	EDCALQMLARSFLEERKEIAKESAODAFLOLLQERHQOQLLSVTALEARHQOALG	917
QY	1154	DKTALTKR---LIIDTVTIGELHELADKAKLVSDHKHPDADQIKQ-LRQOFDTLREKRYE	1209
Db	918	ELTASLESKQGALLAARVAELOTKHAADLGALETRLHSLTLDLSLSECVLSEFQITIRE--	974
QY	1210	SNPVKHYTDMGTFHNKALB---ANVD-----AVKAFINAFKKEHHGVNLTTRT	1254
Db	975	-----HQALELLRADFEQIWKDLSLHOTILTQELEKLRKHKGELQSVRD	1021
QY	1255	VLESQGSALAKKLNLLSLDSGSMFSRSYGGGVSTFVPTLSKKVPVPVPIPGAGIT	1314
Db	1022	HLRTEASTELAGTVAHELQGVHGE-----	1046
QY	1315	LDRAYLNSFRTSGGLNVSFGRDGGVSGNMIVATGHVMPYMTGKTSAGNASDWLSAKH	1374
Db	1047	-----FGSE-----KKT-----ALH	1056
QY	1375	KISPDLRIGAASVGTLOGTFLONSLAKFLTE-----DELPCTHGLTH---GTLTPAELLQ	1426
Db	1057	EKEETLRQSAQAQPPHQBESLSLQLOKKNHQVOOLKQDVLUSLSHETEECHSELEVLQ	1116
QY	1427	KGIEHOMQOGSKLTFSVDTSANLDRAGINL--NEDGSKPNGVTARVSAGLSASANLAAG	1484
Db	1117	QRRERENREGANLLSM-----LKADVNLSHSERGALQDALRLL-----	1155
QY	1485	SRERSTTSQFGSTTSASNNRPTFLNVCAGANL---TAALGVYAHSTHEGKPVGIFPAF	1541
Db	1156	-----GLFGETLRAA---VTLRSRIGERVGLCLDDAGALGALSTA-----PAL	1195
QY	1542	TSTNVSAAL-ALDNRTSQ-----SISLELKRAEPVTSNDISELTSILGKHFK-----	1587
Db	1196	ETWSDVALPELDRTLSECAEMSVAEISSHRESPLMSPESVRECEQPIRRVFQSLSLA	1255
QY	1588	-DSATTKMLAALKELDDAKPAQLHI-LQHPFSAKDVVDGERVEARNLKKLVIROQAAD	1645
Db	1256	VDGLMEWALDSSRQLEEAR---QLHSREKEFSFXN---EETAQVVRKKHQELLECITEES	1309
QY	1646	SHSHELGSASHST--TYNNLSRINDGIVELLKHIFDA-----ALPASSAKR-IGEMMNND	1698
Db	1310	AAKAEALALELHKTOGTLEGF-KVETADUKEVLGAKEDSEHRLVLELESRLRQIQAAEQ	1368

QY	1699	PALKDILIKLOSTPFSASVSMELKDGLRQTETKAILDGKVGREVGVLFDQR---- <td>1754</td>	1754
DG	1369	AALREECTRLWSGEATADAAREAAARKVEDLTKEQSETRKQAE---KDRSALLSQM	1425
QY	1755	RVKSVSVGSVSCKSEGFTPALLLGTNSAAMSERNIGFI-----NFKYGQQDNTPRR-	1808
DB	1426	KILESELEQLSOHRCACQA-----EAVTALEGOVASLDKLNRQRFMDQAAERE	1478
QY	1809	-----FTELEGGAQ-ANPQ	1821
DB	1479	HEREEFQEIQRLGOLROAKPO	1502
RESULT 12			
MAMP2_RAT			
ID	MAMP2_RAT	STANDARD;	PRT; 1861 AA.
AC	P15146;		
DT	01-APR-1990 (Rel. 14, Created)		
DT	01-JUN-1994 (Rel. 29, Last sequence update)		
DT	01-OCT-1994 (Rel. 30, Last annotation update)		
DE	MICROTUBULE-ASSOCIATED PROTEIN 2 (MAP2B) [CONTAINS: MAP2C].		
GN	MAP2.		
OS	Rattus norvegicus (Rat).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
OX	NCBI_TaxID=10116;		
RN	[1]		
RP	SEQUENCE OF 1-1694 AND 1726-1861 FROM N.A.		
RC	STRAIN=Wistar; TISSUE=Brain;		
RX	MEDLINE=90251471; PubMed=2339070;		
RA	Kindler S., Schwanke B., Schulz B., Garner C.C.;		
RT	"Complete cDNA sequence encoding rat high and low molecular weight		
RT	MAP2.";		
RL	Nucleic Acids Res. 18:2822-2822(1990).		
RN	[2]		
RP	SEQUENCE OF 1-1694 AND 1726-1861 FROM N.A.		
RC	STRAIN=Wistar; TISSUE=Brain;		
RX	MEDLINE=91060576; PubMed=2174050;		
RA	Kindler S., Schulz B., Goedert M., Garner C.C.;		
RT	"Molecular structure of microtubule-associated protein 2b and 2c from		
RT	rat brain.";		
RL	J. Biol. Chem. 265:19679-19684(1990).		
RN	[3]		
RP	SEQUENCE OF 1-151; 1515-1694 AND 1726-1861 FROM N.A.		
RX	MEDLINE=9021819; PubMed=2326166;		
RA	Doll T., Papadrikopoulou A., Matus A.;		
RT	"Nucleotide and amino acid sequences of embryonic rat MAP2c.";		
RL	Nucleic Acids Res. 18:361-361(1990).		
RN	[4]		
RP	DISCUSSION OF SEQUENCE.		
RX	MEDLINE=89365159; PubMed=2770869;		
RA	Papadrikopoulou A., Doll T., Tucker R.P., Garner C.C., Matus A.;		
RT	"Embryonic MAP2 lacks the cross-linking sidearm sequences and		
RT	dendritic targeting signal of adult MAP2.";		
RL	Nature 340:650-652(1989).		
RN	[5]		
RP	SEQUENCE OF 1695-1725 FROM N.A.		
RX	MEDLINE=94110302; PubMed=8282767;		
RA	Doll T., Meischner M., Riederer B.M., Honegger P., Matus A.;		
RT	"An isoform of microtubule-associated protein 2 (MAP2) containing		
RT	four repeats of the tubulin-binding motif.";		
RL	J. Cell Sci. 106:1633-640(1993).		
CC	-!- FUNCTION: THE EXACT FUNCTION OF MAP2 IS UNKNOWN BUT MAPS MAY		
CC	STABILIZE THE MICROTUBULES AGAINST DEPOLYMERIZATION. THEY ALSO		
CC	SEEM TO HAVE A STIFFENING EFFECT ON MICROTUBULES.		
CC	-!- ALTERNATIVE PRODUCTS: VARIOUS FORMS OF MAP2 ARE PRODUCED BY		
CC	ALTERNATIVE SPLICING OF THE SAME GENE. MAP2C, THE LOW MOLECULAR		
CC	FORM OF MAP2, LACKS THE CENTRAL DOMAIN OF MAP2A/B.		
CC	-!- DEVELOPMENTAL STAGE: MAP2C IS EXPRESSED DURING EMBRYONIC BRAIN		
CC	DEVELOPMENT AND UNTIL POSTNATAL DAY 10. MAP2B IS EXPRESSED		
CC	THROUGHOUT BRAIN DEVELOPMENT.		
CC	-!- SIMILARITY: CONTAINS 3 OR 4 TAU/MAP REPEATS.		

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; X51842; CAA36135.1; -  
 EMBL; X17682; CAA35667.1; -  
 EMBL; X71487; CAA50588.1; -  
 PIR; S07887; S07887.  
 PIR; S10003; S10003.  
 PIR; A37981; A37981.  
 InterPro: IPR001084; -  
 Pfam: PF00418; tubulin-binding; 4.  
 PROSITE; PS00229; TAU\_MAP\_1; 3.  
 Microtubules; Repeat; Alternative splicing; Calmodulin-binding.  
 DOMAIN 1454 1474 CALMODULIN-BINDING (POTENTIAL).  
 REPEAT 1664 1694 TAU/MAP MOTIF.  
 REPEAT 1695 1725 TAU/MAP MOTIF.  
 REPEAT 1726 1756 TAU/MAP MOTIF.  
 REPEAT 1757 1788 TAU/MAP MOTIF.  
 VARSPLIC 152 1514 MISSING (IN ISOFORM MAP2C).  
 VARSPLIC 1695 1725 MISSING (IN ISOFORM WITH 3 TAU/MAP REPEATS).  
 SEQUENCE 1861 AA; 202409 MW; 42DCF116D21EF54E CRC64;

Query Match 2.1%; Score 194; DB 1; Length 1861;  
 Best Local Similarity 18.9%; Pred. No. 0.15;  
 Matches 407; Conservative 274; Mismatches 819; Indels 652; Gaps 98;

QY 44 LAEECKNRGCKPRIHQPSTAADGISAHQKKSFLRGCLGKTKKFSRSPAGQPGTHSK 103  
 DB 1 MADERKDEKAP--HWTASLTE-AAAHPSPEMKDQGGSG-EGLSRSA-NGEPPYREEE 55  
 QY 104 GATLLDRLARDGGEQHEAAAPDAARLTRSGGVK-----RRNDDMAGR----- 147  
 DB 56 GAF-----GEHGSQGTYS-----TKENGINGELTSADRETAEEVSARIVQVVTAE 102  
 QY 148 -PMVKGSGGDEKVPYQKQRLHNNFGMRQTMLSKMAHPASANAGDRLOHSPHPIGSHH 206  
 DB 103 VAVLKG-----EQEQAQHK-----QPAALPLAAEETVNLPPSPSPASEQ 145  
 QY 207 ELK-EEPVGTSKATTADRV--ETAQEDDDSEFQQLHQRLARENP---PQPP--- 257  
 DB 146 TAALEEDLTASKMEFFEPOOKLPSSFAEPLDKEETEFKMQSKPGEDFEHAALVPQDTSK 205  
 QY 258 -----KLGVATPISARFQPKLTAVAESVLEGTDTTQS-----PLKQSMK 298  
 DB 206 TPQDKKDPQDMGEKSPASPPAQTEGTNLEDI-KQITEPSITVPSIGLSAEPLAPKDQKD 264  
 QY 299 -----GSGAGVT--PLAVTLDKGLQLAP-----DNP-PALNTLLKQTL 334  
 DB 265 WFIEMPVESKKDEWGLAIPISGPLTPMRKQVLEIDPRWEKGQFDSPPSPPHGGSFTL 324  
 QY 335 GKDTQHYLAHASSDQSQ-----HLLDNKGHLFDIKSTATSVSVLHNSHPGEIKG--- 385  
 DB 325 PLDT---VADERVTGSGQFPAPVFFQSDDKMSLDQDTGSGSATSKESKDEEPQDKADKVA 381  
 QY 386 --KLAAGT---GSVSDGKSGKISLG--SCTQSHNKTMLSOPGEAHRSLLTGIWQHP 436  
 DB 382 DVPVSEATVLDGVHSPAVEGVFGENISCEKGTDDQEKKETSTPSVQETLT----- 434  
 QY 437 AGAARQ-----GESTRLHDDKTHILHPELGVWQSAKDKTHSLSRQ 478  
 DB 435 --ETEPQTKLEETSQVSIETVAKEEESLKDQK-----AGVITQSTEQSFREDOK 485  
 QY 479 A-DGLYALKDN-----RTLQNLSDNKSSEKLVDKIKS--YSVDQRCQVAIL--TDTFGRH 529  
 DB 486 GQEQTEIALKQDSFFPISUQAVTDAAMATKTKLEKVTSEPAEVAEVEKQKLEEDIAKDS 545

QY 530 KMSIMPILDASPESHISLSLHFADAHQGL-----LHGKSELAQSVASIGHRLVADSEG 584  
 DB 546 KLE-----GAGSATVAEEMPFYEDKSGMSKYFETSALKEDVTRSTGLGSDGYEUSDSRG 600  
 QY 585 ---KLFSAAIPKQDQGNELKMKAM---POH-----ALDEHFGHDH----- 618  
 DB 601 NAQESLDTVSPKNOODEKELLAKASQSPSPAHEAGYSTLAQSYTSDHPSELPEEPSSPOE 660  
 QY 619 ---QISGFFHDDHQLNALVNNFRQOACPLGND---HOFHPGNML---TDALVIDNQ 668  
 DB 661 RMFTIDPKVYGEKRDLSKKNKDDLTLSRSLGLGRSAIEQRSMISNLPMSCLDSIALGFN 720  
 QY 669 LGL-HHTNPPEHEILDMGHLSLALQEG-----KLHYF----- 700  
 DB 721 FGRGHDLSPLASDILTN---TSGSMDEGDDYLPPTTPAVEKIPCFPIESKEEEDKTEQAK 777  
 QY 701 ---DOLTKGWTGAESD-----CKOLKKGDLGA-----AYLLKDGVEKRLNI 738  
 DB 778 VTGGQTTQVETSSSEPPFAKEYYKNGTVMAPDLPEMLDLAGTRSLASVSADAEEVAR--- 834  
 QY 739 NQSTSSIKHGTEVFSPLPHVRNKPEPGDALQGLNKDDKAQAMAVICVKNKYALTEKDIR 798  
 DB 835 RKSVPSEAVAESSTGLPPVADDSQP-----VKPDSQLEDGMYCVFNKY----- 878  
 QY 799 SFQIKPGTQOLERPAQTLRSREGISGE-----LKDIIHVDHKQNLALTHEGEV- 845  
 DB 879 -----TVLPSPVQ---DSENLGSGSFYEGTDDKVRDRDLATDLSLIEVLAAGRKY 929  
 QY 846 --FHOPREAW-ONGAESSSWHK-----LALPOSESKLSLSDMSHEHKPI--- 886  
 DB 930 DEPTAEKEASPPSADKSGLSREFQDRKANDKLDTVLEKSEHVDSKEHAKSESEVGDK 989  
 QY 887 -----ATPEDGSOHOLKAGWHAYAAPERGFLAVGTSGSQTVFNRLMOG--VKCKVIP 937  
 DB 990 VELFGLGVITYEQTSAKELITTKETAPERAEKGLSSVPEVAEVETTTKADQGLDVAARKDD 1049  
 QY 938 GSGITVKLS---AOTGGMTGAERKRVSSKFSERIRAYAFNPTMTSPRIKNAAYATQHGW 994  
 DB 1050 QSPLDIKVSDFGQMASGMSVDAGTKIELKFVDQQL-----TLSSPAQETDSFM----- 1099  
 QY 995 QGRGLPLYEMOGALIKOLDAHNVYRNAPQPDLOSKLETLD-----LGEHGAELLND 1047  
 DB 1100 ---GISSHVKDGAKVSETE---VKEKVAKPDLVHQ-EAVDKESYESGEHESITMES 1151  
 QY 1048 MKRFDELEQASRSTVTLGOHGVLK-----SNGEINSEKPSPG 1088  
 DB 1152 LKPDGKKE---TSPETSLIOEVALKLSVEIPCPPPPSEADSSIDEKAEVOMEFTQLPK 1208  
 QY 1089 KALVQSFNVNRSQDLSKSLQQAQVHATPPSAESKLSQSLGHEFVSAGVDMSHQKGEIPLGR 1148  
 DB 1209 EESTETPDIPALPSDVTQPOPEAVVSEPAEVRGEBEEI-----EAEGEY--- 1252  
 QY 1149 QRDNDKALTAKSL-ILDTVTIGELHELADKAKLVSHDKPADQLKQLRQOFTDLUREKR 1207  
 DB 1253 ---DKLLFRSDTLQITDILLVPGSREEFVETCP--GEHKGVVESVVTIEDDFITV--- 1301  
 QY 1208 YESNPVKHYTDMG-----FTHNKALE----- 1228  
 DB 1302 ---VQTTDEGELGSHSVRFAPVQPEERRPDPHDEEVLMAAEAAQAEKQSGPDA 1356  
 QY 1229 -ANYDAVKAFINAFKKEHHGVNLTTRTV-----LESQSGA 1262  
 DB 1357 PATPEKEEVPFSEYKTEYDDYKDETTIDDSIMDADSLMVDVTDQDDRSILTEQLETIPKE 1416  
 QY 1263 ELAKKLNLTLSLDSGESMSFSRSGGVSTVFTLSKKVPVPGAGITLDRAYNLS 1322  
 DB 1417 ERAEK-EARRPSLEKHKRKEKPFKTRGRISTPERREVAKEKPESTV----- 1460  
 QY 1323 FSRTSGGLNVSGDGGVSGNIMVATGHDVMPYMTGKTKTSAGNASDWLSAKHKISPDRLI 1382  
 DB 1461 -SRD-----EVRKKAVYKKAELAKESEVQAHSPSRLILKPAIKYTRPTH--LSCVKRK 1512  
 QY 1383 GAAVSGTLQGTQLQNSLKFKLTEDELPGFIHGLTHGLTLPALLQKGIHQMGSKLTFS 1442

Db 1513 TTATSGE---SAQAPSAFQAKDKV---TDGIT-----KSPKRSLL 1548  
QY 1443 VDTANLDRAGINLNEGDSKPGVTARYSAGLSASANLAAGSRERSTTSGQFGSTTSAS 1502  
Db 1549 PRPSILPRRCVSGDREN-----SFLNSSISSARRTRSEPIPRACKSG 1595  
QY 1503 NNRPFTLVGAGANLTAALGVAHSTHEGKP-VGIFPAFTSTNVSAALDNRNTSOSIS 1561  
Db 1596 TSTPT---TPGSTAITPGTPPSYSRTPTGTPSPVPTGPKSGILVPSEK-----K 1646  
QY 1562 LELKRAEPTVSTNDISELSTGLKHKFSDATTKMALKELDKAPBQHLQHFSAKD 1621  
Db 1647 VAIIRTPP-----KSPATPKQLRLINQ-----PLPDLKNVSKSGISTD 1694  
QY 1622 VVGDERYE-----AVRNLLKLVIRQAADSHSMELGSA-SHSTTYNNLSRINDQIVVELLH 1676  
Db 1685 NI---KYQPKGQVRILNK-----KMDFSKVQSGCKNDIKHAGSGNVQIVT 1730  
QY 1677 KHFDALPASSAKRLGEMNNDPALKDIKQLQSTPFSSASVSMELKDLREQTEKAILD 1736  
Db 1731 KKID-----LSHVTSCGSLKNIRHRPGGRVKIESVKLDFKE--KAQAKVGSGLD 1778  
QY 1737 -----GKVGREEVGVLFDQRNLRVK--SVSVSQSVSKSGFNTPALLLGTSTNSAAM 1786  
Db 1779 NAHVPGGKVKIDSQKLFREHAKARVDHGAETITQSPSRS-SVASPRRLSNVSSS-- 1834  
QY 1787 SMERNIGTFNKYGQDTPRFTLEGGIAQANQVASALTD---LKKEGL 1834  
Db 1835 -----GSINL-----LE-----SPQLATLAEDVTAALAKOGL 1861

## RESULT 13

PGCV\_MOUSE STANDARD; PRT: 3358 AA.  
AC Q62059; 062058;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE VERSICAN CORE PROTEIN PRECURSOR (LARGE FIBROBLAST PROTEOGLYCAN)  
DE (CHONDROITIN SULFATE PROTEOGLYCAN CORE PROTEIN 2) (PG-M).  
GN CSPG2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
[1]  
RP SEQUENCE FROM N.A. (VARIANTS V0; V1 AND V2).  
RC STRAIN-C57BL/6, AND SWISS WEBSTER; TISSUE=Brain;  
RX MEDLINE=95122551; PubMed=7822336;  
RA Ito K., Shinomura T., Zako M., Ujita M., Kimata K.;  
RT "Multiple forms of mouse PG-M, a large chondroitin sulfate  
proteoglycan generated by alternative splicing.";  
RL J. Biol. Chem. 270:958-965(1995).  
RN [2]  
RP SEQUENCE OF 1-348 AND 3053-3358 FROM N.A. (VARIANT V3).  
RC STRAIN=C57BL/6;  
RX MEDLINE=95181355; PubMed=7876137;  
RA Zako M., Shinomura T., Ujita M., Ito K., Kimata K.;  
RT "Expression of PG-M(V3), an alternatively spliced form of PG-M  
without a chondroitin sulfate attachment in region in mouse and human  
tissues.";  
RL J. Biol. Chem. 270:3914-3918(1995).  
CC -!- FUNCTION: MAY PLAY A ROLE IN INTERCELLULAR SIGNALING AND IN  
CONNECTING CELLS WITH THE EXTRACELLULAR MATRIX. MAY TAKE PART IN  
THE REGULATION OF CELL MOTILITY, GROWTH AND DIFFERENTIATION. BINDS  
HYALURONIC ACID  
CC -!- SUBCELLULAR LOCATION: SECRETED; EXTRACELLULAR MATRIX.  
CC -!- ALTERNATIVE PRODUCTS: AT LEAST 4 ISOFORMS; V0 (SHOWN HERE), V1,  
V2 AND V3; ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC -!- TISSUE SPECIFICITY: V2 IS FOUND ONLY IN BRAIN.  
CC -!- DEVELOPMENTAL STAGE: DISAPPEARS AFTER THE CARTILAGE DEVELOPMENT.  
CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.

CC -!- SIMILARITY: CONTAINS 2 LINK DOMAINS.  
CC -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.  
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 SUSHI (SCR) REPEAT.  
CC -!- SIMILARITY: BELONGS TO THE AGGREGAN/VERSICAN PROTEOGLYCAN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; D16263; BAA03796.1; -  
DR EMBL; D28599; -; NOT\_ANNOTATED\_CDS.  
DR EMBL; D32040; BAA06802.1; -  
DR HSSP; F00740; IIXA.  
DR MGD; MGI:102889; Cspg2.  
DR InterPro; IPR000152; -  
DR InterPro; IPR000436; -  
DR InterPro; IPR000538; -  
DR InterPro; IPR000561; -  
DR InterPro; IPR001304; -  
DR InterPro; IPR001438; -  
DR InterPro; IPR001881; -  
DR InterPro; IPR003006; -  
DR Pfam; PF00008; EGF\_2.  
DR Pfam; PF00193; Xlink; 2.  
DR Pfam; PF00047; Ig; 1.  
DR Pfam; PF00059; lectin\_c; 1.  
DR Pfam; PF00084; sushi; 1.  
DR PRINTS; PRO0010; EGFBL00D.  
DR PROSITE; PS00010; ASX\_HYDROXYL; 1.  
DR PROSITE; PS00022; EGF\_1; 2.  
DR PROSITE; PS01186; EGF\_2; 1.  
DR PROSITE; PS01187; EGF\_CA; 1.  
DR PROSITE; PS01241; LINK; 2.  
DR PROSITE; PS00615; C-TYPE LECTIN\_1; 1.  
DR PROSITE; PS00411; C-TYPE LECTIN\_2; 1.  
KW Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;  
Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;  
Hyaluronic acid; Alternative splicing.  
FT SIGNAL 1 20 POTENTIAL.  
FT CHAIN 21 3358 VERSICAN CORE PROTEIN.  
FT DOMAIN 37 137 IG-LIKE V-TYPE DOMAIN.  
FT DOMAIN 167 244 LINK 1.  
FT DOMAIN 265 334 LINK 2.  
FT DOMAIN 348 1308 GAG-ALPHA (GLUCOSAMINOGLYCAN ATTACHMENT  
DOMAIN).  
FT DOMAIN 1309 3052 GAG-BETA.  
FT DOMAIN 3052 3088 EGF-LIKE 1.  
FT DOMAIN 3090 3126 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 3129 3255 C-TYPE LECTIN.  
FT DOMAIN 3258 3316 SUSHI.  
FT DISULFID 44 130 BY SIMILARITY.  
FT DISULFID 172 243 BY SIMILARITY.  
FT DISULFID 196 217 BY SIMILARITY.  
FT DISULFID 270 333 BY SIMILARITY.  
FT DISULFID 294 315 BY SIMILARITY.  
FT DISULFID 3056 3067 BY SIMILARITY.  
FT DISULFID 3081 3076 BY SIMILARITY.  
FT DISULFID 3078 3087 BY SIMILARITY.  
FT DISULFID 3094 3105 BY SIMILARITY.  
FT DISULFID 3099 3114 BY SIMILARITY.  
FT DISULFID 3116 3125 BY SIMILARITY.  
FT DISULFID 3132 3143 BY SIMILARITY.  
FT DISULFID 3160 3252 BY SIMILARITY.  
FT DISULFID 3228 3244 BY SIMILARITY.  
FT DISULFID 3259 3302 BY SIMILARITY.  
FT DISULFID 3288 3315 BY SIMILARITY.  
FT CARBOHYD 57 57 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CARBOHYD 330 330 N-LINKED (GLCNAC... ) (POTENTIAL).

```

FT CARBOHYD 351 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT CARBOHYD 441 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT CARBOHYD 807 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT CARBOHYD 914 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT CARBOHYD 951 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT CARBOHYD 1305 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT CARBOHYD 1372 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT CARBOHYD 1679 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT CARBOHYD 1729 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT CARBOHYD 2054 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT CARBOHYD 2244 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT CARBOHYD 2362 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT CARBOHYD 2627 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT CARBOHYD 3030 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT CARBOHYD 3332 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT CARBOHYD 3342 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT VARSPLIC 349 MISSING (IN ISOFORM V1) .
FT VARSPLIC 1309 MISSING (IN ISOFORM V2) .
FT VARSPLIC 349 MISSING (IN ISOFORM V3) .
FT CONFLICT 348 P -> R (IN REF. 2) .
SQ SEQUENCE 3358 AA; 366938 MW; 071B80026BC0762D CRC64;

Query Match 2.1%; Score 194; DB 1; Length 3358;
Best Local Similarity 18.1%; Pred. No. 0.36;
Matches 388; Conservative 287; Mismatches 791; Indels 680; Gaps 104;

Qy 25 GVALQOQSSSS---PQNAASLAEGKNGRK-----MPRIHQPSTAAD-GIS 68
Dy 724 GMLPTSSSESVESVSPALTIE-KLTGKPTAEARDEVEETLRLTLETDVTKSKDVT 782
Qy 69 AAHQCKKSFSLGCL-----GFKKFSRPAQOGPGTHSKGATLRLLARDGCEQHEAA 123
Dy 783 RVHLTHSLNVEVTVVSKWPGDEDNSTKPLPSTEAGTKLPVPPLSTGINKDKRE 840
Qy 124 APDAARLFRSG-----YKRRNMDMAGRMV-----KGGSGE-- 156
Dy 841 IPS---FTDGGGEYTLFPDGTPKLKVSEEDLASGELTVFHTSTSGSAEKSASGEPT 897
Qy 157 --DK-VPTQORHQLNFGQMTLSKMAHPASANAGBRLQHSPPHPCSH-HEIKEEP 212
Dy 898 TGDRLPLTSTEDQVIN-----ATAEGSALGEDTEASKPLFTGPPFVHTSDVEELAFVN 951
Qy 213 VGSTSKATT-----AHADRV-----ETAODDDSEFQQLHQORLAR- 248
Dy 952 YSSTQEPYVDISHTSPLSLIPKTEWSVLETSPVLEDEILCKSDQDILEQTHLEATMSP 1011
Qy 249 -----ERENPPQPKLGVA TPISARFQPKLTAVAESVLEGTDTTQSPKLPQSM 297
Dy 1012 GALRTTGVSGQETOEPQTPGSPFTFSSTAVMAKETTAPE-----EGEGSTYTPSEGR-LM 1067
Qy 298 KGS-----GAGVTPPLAVTLDKGL---OLAPDNPPALNTLLKQTLCKDQTOHYLAHASSDG 350
Dy 1068 TGSEKVPGLTTPVGTSPGAIQDOEVEMDTMTVLMSTIRPTVVSSTSEVIYEA--EG 1125
Qy 351 SQHLLLDNKGHLFDIKSTATSVSLHNSHPGEIKGLAQAGTGSVDGKSGKISLCSGT 410
Dy 1126 SSPT-----ERASTLRPE-----QTHVTQLMEETTEBGR-KASLD-----YTDLGSGL 1167
Qy 411 QSHNKTML-----SQPEAHRSLLTGIWQ-HPAGAAAPQ----- 443
Dy 1168 FEPRA TELPKPSTPSDI--SVFTAIDSLHRTPLSPSSSFTTEEQRVFEESSEKTTGDI 1225
Qy 444 --GESIRLH-----DKIHI-----LHPELGVWQSAKDQTHSLSRQ 478
Dy 1226 LPGESVTOHPVTLIDIVAMKTESIDHMTSKPPVTPQTRPSVVERKTTSKTQELSTSTP 1285
Qy 479 ADGKLYALKDNRTLQNLSDNKSSEKSLVDKIKS-YSDV--QRGQVAILTTPGGRHKM--SI 533
Dy 1286 AAGTKFHDINDIYIIEVRENKTRGLSDMIVSGHPIDSEKSEEPCESEETDPLHDLFAEI 1345
Qy 534 MPSLDASPESHLSLSLHAFADAHQGLLHGKSELEAOSVAISHGRVLVADSEGLFSAAPK 593
Dy 1346 LPPELPDSFEIDIYHSEEDGEDCVCNADTVTTTSPVQVINGKQLV-----TTVPK 1396

```

```

Qy 594 QGDCNELKMKAMPOHALDEHFGHDHQISGFFHDDHQLNALVKNFNPRO----- 642
Dy 1397 DPEAAEARQGVESVAPSQNFPSAT-----DTHQFILAATESSTMOFKKSKEGTELL 1451
Qy 643 -----HACPLGNH-----QFHPG-----W--NLTDALV-IDNOLGLHHTNPE 677
Dy 1452 EITWKPEYPTPDHVSSEGPDPVFTLSHOGKTRTWSESITTESPNLENPV---HKQPK 1508
Qy 678 PHEILDMHGLSLALQEGKLH-YFDQLTKGWTGAESDCKQLKKGLDGAAYLLKDGVEKRL 736
Dy 1509 PVPFLPESSCEGAIEQASQETILSRATEVALGKETD-----Q 1546
Qy 737 NINGSTSSIKGHTENVFSLPHVRNKPPEGDALQGLNKDDKAQAMAVIGNKYLALTEKGD 796
Dy 1547 SPTLSSTILSSSVNVYL-----EEEP-LTLTGISQTDSEMSST-----IESWVEIT--- 1592
Qy 797 IRSFOINPGTQOLERPAOTLS-----REGISGELKDIHVDHKONLYALTHEGEVPH 847
Dy 1593 -----PSQTVKFESSSAPIEG-SGEVE-----NKNKIFN 1623
Qy 848 QPREAWQNGAESSWHKLALPQSE--SKLKS LDM-----HEKHP--IATFEDGSQHQ 896
Dy 1624 MVTD-----LPQRDPTDTLSP LDM SKIMITNHHIYIPATIALDS----- 1663
Qy 897 LKAGWHAYAAPERGPPLAY--GTSQSQTIVNRLMOGVKGVIPGSLTVKLSAQTGCMTG 954
Dy 1664 -----KLPSDARPTTVWNSNSTSEWSDKSFGRKK-----ENEDEEGAVNA 1707
Qy 955 AEGKVSXKFSERIRAYAFNPTMTSPRIKNAAYATQHGMOG-----REGLKPLVEMQ 1007
Dy 1708 AHQGEVRAA--TERSDHLLLTPELESSNVDSADSLAT---WEGFILETTTTESEKEMANST 1763
Qy 1008 GALIKQLDAHNV-----RHNAPDPLDQSKLETL-----DLG-----EHGAEL--- 1044
Dy 1764 PVFRETIGVANVEAQPFHSSSHSRVQBELTTLGSPNPSLTDLGDGSDASTGMELITAS 1823
Qy 1045 -----LNDMKFRFDELEOSATRSVTVLGOHOGVLKSNGEINSEFPK---SPGKALVQSFN 1096
Dy 1824 LFTLDLSETKVKKELPSTPSPV-----EISSFEPTGUTPSTVL-----D 1865
Qy 1097 VNRSG---QDLKSKLQQAHVATPSPAESKLSQMLGHFVSAGVDMSHQKEIPLGRORDPN 1153
Dy 1866 IEIAGVMSQTSOKTLISEISGKPTS-QSGVRDLYTGF-PMGEDFSGDFEYPTVSYPTMK 1923
Qy 1154 DKT---ALTKSRLLDVTIGELHELADKAKLVSDHK-PDA-----DOIQLQRLQ 1198
Dy 1924 EETVCMGSGDDERVRDTQTSSTIPTSDNIYVPDPSKGPDPSTVASTTAPFWEEMSSAEG 1983
Qy 1199 QFDTLREKRYESNPVKHYTDMCF--THNKALEANYDAVKAFINAFKKEHHGVNLTTRTVL 1256
Dy 1984 SGEGLASVRSSGVPLPLAVDIFSGTESPYFDEEFEEVAATVTEANERP-----TVL 2034
Qy 1257 ESOGSAELAKKLKNTLLSLDSEMSFSRSYGGVSTVFVPTLSKKVPVPV---IPGAG 1312
Dy 2035 PTAAGNVTDLTENCYIEVNSTMSLDFPQ-----THEPSKLWSKPEVNLQKEIGRET 2087
Qy 1313 ITLDRAYNLSPSRTSGGLNVNFGRDGV-----SGNIMVATGHVMPYMTGK 1359
Dy 2088 VTKEAQQ--QKTPESLHSSFAPEQTILETOSLIETEFQTSYDYSMLTT---LKTITNK 2141
Qy 1360 KTSAGNASDWLSAKHKISPDLRIGAAVSGTLOGLTQNSLKFKLTEDELPGFHLGHTGL 1419
Dy 2142 EVE-----EEGSIAHMTSP-----GPGIKDLESYTHTEPAFCKSHSFATAL 2184
Qy 1420 TPAELLQKIEHQMKQSGKLTFSVDTSANLDRAGTINLEDGSKPVGTVARVARSAGLSASA 1479
Dy 2185 V-----TESGAARSVLMDSSSQEE---ESIKLFOKGVK----- 2214
Qy 1480 NLAAGSRERSTTSQFGSTTSASNRRPTFLNGVAGAGANLTAALGAHSHSTHEGKPVGIFP 1539
Dy 2215 -----LTKNESADLSFSGLSGSG-----GALP 2236

```

Qy 1540 AFTSTNVS-----AALALDNRSTQSISLEL-----KRAEPTVSTNDISELTSLG 1583  
 Db 2237 PLPTTSVNLTKMIISTLYAETHSMESLGTSLGDKMEDHEDYSSNEVRLISKIG 2296  
 Qy 1584 KHKDSA-----TTKMLAAKELDDAKPAEQLHILQOHFSKADVVGDERY 1628  
 Db 2297 SISQDSTEALDITLSTHTGTEPTTSLPFVKMLDLSRPQ-----DPSGGKRK 2345  
 Qy 1629 EAVNLKLL--VIRQAAADSHMELGASHST-----TYNNLSRINNDGIVELLKHEDAA 1682  
 Db 2346 KXTHRPQTMGLSINENSASAEAGATSPATFLPQIYS-----VEMT-KHF--- 2391  
 Qy 1683 LPASSAKRLGEMMN-----NDPALKDIKIQSQSTPFSSASVSMELKDLGRLREOTERAIL 1735  
 Db 2392 --APSESQPSDLFNVNSGEGSEVDLVLVYTSQTQASSQGDMSLASHGFLEK----- 2443  
 Qy 1736 DKVGREEVG-----VLFQDRNLNRKVSVSQSVSKSEGFNPALLLGTSNSAAM 1786  
 Db 2444 HPEVSKTEAGATDVSPASAMFLHSEYK--SSLYPTSTLPTSTPEYKSP-----SEGIED 2496  
 Qy 1787 SMERNIGTINFKYGQDQNTPRFTLEGGAQANPQVASALTDLKE 1832  
 Db 2497 GLQDN---IQFE--GSTLKFPSRRKTE-----SIIDLDKE 2527

## RESULT 14

P3P\_LACLC STANDARD; PRT; 1902 AA.  
 AC P15292;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE PIII-TYPE PROTEINASE PRECURSOR (EC 3.4.21.96) (LACTOCEPIN) (CELL WALL-  
 DE ASSOCIATED SERINE PROTEINASE).  
 GN PRTP.  
 OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).  
 OG Plasmid.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Lactococcus.  
 ON NCBI\_TaxID=1359;  
 RX MEDLINE=89340435; PubMed=2760036;  
 RT Vos P., Simons G., Slezien R.J., de Vos W.M.;  
 RT "Primary structure and organization of the gene for a procaryotic,  
 RT cell envelope-located serine proteinase.";  
 RL J. Biol. Chem. 264:13579-13585(1989).  
 CC -!- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE  
 CC GROWTH OF THE BACTERIA ON MILK.  
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE  
 CC SUBTILASE FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; J04962; AAA03533.1; ALT\_SEQ.  
 DR PIR; A32634; A32634.  
 DR HSSP; P00782; 2SBT.  
 DR MEROPS; S08.019; -.  
 DR InterPro; IPR000209; -.  
 DR InterPro; IPR001899; -.  
 DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
 DR Pfam; PF00082; Peptidase\_S8; 3.  
 DR PRINTS; PR00723; SUBTILISIN.  
 DR PROSITE; PS00136; SUBTILASE ASP; 1.  
 DR PROSITE; PS00137; SUBTILASE HIS; 1.

DR PROSITE; PS00138; SUBTILASE\_SER; 1.  
 DR PROSITE; PS00343; GRAM\_POS\_ANCHORING; 1.  
 KW Hydrolase; Serine protease; Cell wall; zymogen; Signal; Plasmid;  
 KW Transmembrane.  
 FT SIGNAL 1 33  
 FT PROPEP 34 187  
 FT CHAIN 188 1902  
 FT DOMAIN 188 1876  
 FT TRANSMEM 1877 1895  
 FT DOMAIN 1896 1902  
 FT ACT\_SITE 217 217  
 FT ACT\_SITE 281 281  
 FT ACT\_SITE 620 620  
 FT DOMAIN 1867 1872  
 FT SEQUENCE 1902 AA; 200550 MW; 87CECBAA9345F9D3 CRC64;

Query Match 2.0%; Score 193.5; DB 1; Length 1902;  
 Best Local Similarity 18.9%; Pred. No. 0.16;  
 Matches 409; Conservative 284; Mismatches 720; Indels 753; Gaps 113;

Qy 11 KAAVH-----TAAHNPVGHGVALQOG-----SSSSPQNAASAAAEKGNKGRKPRHQPS 62  
 Db 32 KAAISQQTGGSLANTVTAATAQAATDTTAATTQAATQAAKIDYKLNKLVQOQDI 91  
 Qy 63 AADGISAAHQKKFSLRGCLGTRKFSRPAQOGPG---TTHSKGATLRDLLARDDETQ 119  
 Db 92 YVDIVVQ-----MSAAPASENGILRTDYSSTA-----EIQ 121  
 Qy 120 HEAAAPDAARLTRGGVKRRNMDMAGR-----MVKGSGEDKVPQCKRHQNNFQOMQ 176  
 Db 122 QETNKVIAQAQSKAAVEQVT-QOTAGESYGVVNGFSTKVRVVDIPK---LKQIAGVKT 177  
 Qy 177 TMLSKMAHPASANA-----GDRLOHSPHPGSHHEIK---EPVGS 215  
 Db 178 VTLLKVIPTDANKSMANVQAVSNYKKGEGTVVSVIDSGIDTPHKMRLSDDDKDKVL 237  
 Qy 216 TSKATTAHADRV-----IAQEDD---DSEFQQLHQORLA----- 247  
 Db 238 TKSDEKFTVTKHGRVFNKVPYGFNYADNNDTITDDKVDQHGHHVAGIANGCTGDD 297  
 Qy 248 -----RENNPPPKLGVATPLISA----- 267  
 Db 298 PAKSVVGVAPAEQALLAMKVFNSDTSAKTSATGVVSAIEDSAKIGADVNLMSLGSNGNQ 357  
 Qy 268 -RFPQKLTV-----AESVLEGTDTTQSPKLPQSM---KGSGAGVT 305  
 Db 358 TLEDPELAAYQANESGTAAVISAGNSGTSGSATEGVNKKDYGLQDNEMVSGFTSGAT 417  
 Qy 306 PL-----AVTLDKGK-LQLAPDNPALNTLLKOTLGKDTQHYLAHHASSDGSQ- 352  
 Db 418 TVSAENTDVTQAVTITDGTGLQLGP-----ETQLSSHDFTGSFQDK 461  
 Qy 353 --HLLDNKCHLFDIKTATSYSVLHNSHPGEIKKLAQAGTGSVSD-----GKS 401  
 Db 462 KFIYVDKASGNL--SKGALADYT-----ADAKGKIAIVKRGESFDDKQYAAAGAA 512  
 Qy 402 GKISL---GSGTQSHNKTMLSPGFAHRSLLTG-----IWQHPACAARPGQESIRL--- 449  
 Db 513 GLIIVNTDGTATPMTSIALITTTFTFGLSVYTGKLVDMVTAHPDDSL---GVKITLAML 569  
 Qy 450 -----HDDKI-----HILHPELGVWQSAKDTHDSOLS----- 476  
 Db 570 PNOKYTEDKMSDFTSYGPFVSNLSFKPDITAPGNIWSTQNNNGYTNMSTGSMASPTIAGS 629  
 Qy 477 -----RQA-----DGKLYA-----LKDNTLQ 494  
 Db 630 QALLKQALNNKNNPFYAYYKQKLTALTDFLKTVEMNTAQINDINNNVIVSPRQAG 689  
 Qy 495 LSNKSKSEKLVDKIKSYSVQDQGOVAI-----LTDTPGRRHMS-----IMPSLDASPE 542  
 Db 690 LVDVKAIDALEKRNPSVVAENGYPAVELKDFSTDKTKFLFTNRTTTHLTQVQMSNTD 749

Qy 543 SHISLSLHFAHQGLLHCKSELAQSVASHGRVLVADSEKL--FSAAIKPGDGNELK 601  
 Db 750 TN-AVYTSATDPSNGLVYDK-KIDGAALKAASNTVPAGKTAQIEFTLSLPSFDOQOF- 806  
 Qy 602 MKAMPOHALDHFGHDHQSIFGFFHDDHGLNVALVKNRQHQHACPLGNDHQ-FHPCWNL 660  
 Db 807 -----VEGFL-----NFKGSGSLNLPYMGFFGDWN-- 833  
 Qy 661 DALVIDNOLGLHHTNPPEHILDMHGLSLALQECK--LHVFDOLTKGWTCAE----- 711  
 Db 834 DGIIVDSLNGITYSPAG-----GNFGTVPLKNNKNTCTQYGGMTVDADGNKVDDQA 886  
 Qy 712 ----SDCKOLKGLDGAAYLLK-----DGEVKRLNINOSTSSIKHGHTENFSLPH 757  
 Db 887 IAFSSDKNALYNDISMKYLLRNISNVQVDILDGQGNKVTTLSSSTNRKKTYYNAH500Y 946  
 Qy 758 V-RNKPEPDALOGGL-----NKDDKAQAMAVGVNKKYLALETEKDIRSFQIKPGT 806  
 Db 947 IYVYAP-----AWDGTYYDQDGNIKTADDSYTYRISGV-----PEGGDKR--QVFDVP 994  
 Qy 807 QOLERAQTLRSIGSELKD-----IHVDHKONLYAL-----THEGEVFHOPREAW 853  
 Db 995 FKLDKAPTVRHVVALSAKTENGKTQYLLTAEAKKDDLSGLDASKVKTETNEVTNLDATFT 1054  
 Qy 854 QNGAESSSMHKLALPOSESKLSL-----DMSHE-----HKPIATFEDGSQHLKAGGWHAY 905  
 Db 1055 DAGTTADGTYKIETPLSDEQAQAGNCDNSAELYLTDNASNATDODASVQ--KPGSTSF 1112  
 Qy 906 AAPERPLAVGTSQVTFNRLMOGVKGVIPGSLGTLVKLSAOTGTMGAEGRKVS5KFS 965  
 Db 1113 LIVNGGIPDKISSTTGYEANTQG--GGYTFSG--TYPAAVDGTYTDAOGKKHDLNTT 1168  
 Qy 966 ERIRAVAFNPTMTPTPRIKNAAYATQHGQWGREGLKPLY--EMOGLIKOLDANVHRNA 1023  
 Db 1169 YDAATNSFTASM--PVTNADYAAQ-----VDLYADKAHTQLLKHFDT-KVRLMA 1214  
 Qy 1024 PQPDQSKLETLDLGEHGAELNDMKFRDELOSATRSVTYVLOGHGVLSKNGEINSEF 1083  
 Db 1215 P-----TETDLK-FNNGSDQTS-----BATIKVGTVSADT 1244  
 Qy 1084 KPSGKALGVNFNRSQDLSKLOQAVHATPPSAESKLQSLMGLHFVSAGVDM5HOKGE 1143  
 Db 1245 K-----TVN-----VGHTVAALDAQHFSVD 1265  
 Qy 1144 IPLGRQRPNDKTALTKSRILDTVTIGELHELADKAKLVSDHDKPDADQIKLRQO--FD 1201  
 Db 1266 VPV---NYGDNT-----IKVTATDKDGNITTEQKTIITSSYDPD-----MLAKSVTF 1309  
 Qy 1202 TLREKRYESNVKHYTDMGFTHNKALEANYDAVKAFLNAFKKEHGVNLTTRTVLESQ-G 1260  
 Db 1310 -----QGVKFGTNKF-----NATSAKFYDPKTG 1332  
 Qy 1261 SAEALAKILKN--TLLSLDS-----GESMSFSRYCGGVSTVFVPTLSKKVPVPVPGAGI 1313  
 Db 1333 IATIGVKVHPTTFLQVNDGKOIPIKDDLTFS-----FTLDGLTGLQK-PFGVVVG--- 1381  
 Qy 1314 TLDRAYNLSFRTSGGLNVSRGSGVSGNIMVATGHVDMFY-----MTGKTSAGNA 1366  
 Db 1382 -DTONKTFQEA-----LSFILD-AVAPTLSLDSSTDPAYTNDPNFQITG--TATDNA 1431  
 Qy 1367 SDWLSAKHIKISPDRLI-CAAVSGTLOGTLQNSLKFKLTEDELPGFIHGLTHGLT-TPAEL 1424  
 Db 1432 -OYLS-----LSINGSVASQYEDININSK-----PG-----HMAIDQPVKL 1468  
 Qy 1425 LO-KGIEHQMGKSLTFSVDTSANLDRAGINLNEDESKPNGVTVARVSAGLSANLAA 1483  
 Db 1469 LEGNV-----LTVAVTDS-----ED-----NTTTKNITVYVPEPKTLAA 1503  
 Qy 1484 GSREHSTT-----SGOFGSTTS-ASNNRPTFLNGVCAGANLTAALGVAH5STHE 1531  
 Db 1504 PVTPTTPEAQTVTITANAATGETVQYSADGGKTYQDVPAGVTITA-----NGTFK 1557

Qy 1532 GKPVGIF---PA--FTSTNVSAALALDNRTSQSISLELKRAEPVTSNDI--SELTSTLG 1583  
 Db 1558 FKSTDLYGNE5PAVDYVVTNIKA-----DPAQLQAQAKQELTNLIASAKTSLASG 1607  
 Qy 1584 KHFKDSATTMKLAALKELDDAKPAEQHLHQHFSKADVVGDDE--YEAVRNI----- 1634  
 Db 1608 KY--DDAITTALAA-----ATOKAQATLDDQTNASVDSLTGANRDLQTAIQLAAKLPAD 1659  
 Qy 1635 KKLVI---RQQAADSHMELSGASHTTNNLSRINDGIVEL-LKHFDAAALPASSAKR 1690  
 Db 1660 KRTSLNQLQSVKDALGTGNOTDPSTGKTFTAALDDLVAAQAGTQTDQLOATLAKI 1719  
 Qy 1691 LGE MNNDPALKDIKLOLQSTPSSASVSMELKDLGRLQTEKA-----ILDGKVGREVG 1745  
 Db 1720 LDEVL---AKLAEGIK--AATP---AEVG-NAKDAATGKTWYADIADTLTSGOASADSD 1770  
 Qy 1746 VL--FODRNNLRVK---SVSVSOSVSKSEGFNTPALLLGTSNSAAMSERNIGTINPKYG 1800  
 Db 1771 KLAHLQALQSLKTKVAABAEAAKTVGKDG-----TTGTSDK-----G 1808  
 Qy 1801 QDQNT 1806  
 Db 1809 GGQGT 1814

RESULT 15  
 P2P\_LACLA STANDARD; PRT; 1902 AA.  
 ID P2P\_LACLA AC P15293;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE P11-TYPE PROTEINASE PRECURSOR (EC 3.4.21.96) (LACTOCEPIN) (CELL WALL-  
 DE ASSOCIATED SERINE PROTEINASE) (LP151).  
 GN PT.  
 OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).  
 OG Plasmid pLP763.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Lactococcus.  
 OX NCBI\_TaxID=1360;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NCDO 763;  
 RX MEDLINE=89313288; PubMed=2501630;  
 RA Kwaki M., Ikemura H., Shimizu-Kadota M., Hirashima A.;  
 RT "Molecular characterization of a cell wall-associated proteinase gene  
 RT from Streptococcus lactis NCDO763.";  
 RL Mol. Microbiol. 3:359-369(1989).  
 CC -!- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE  
 CC GROWTH OF THE BACTERIA ON MILK.  
 CC -!- CATALYTIC ACTIVITY: ENDOPEPTIDASE ACTIVITY WITH VERY BROAD  
 CC SPECIFICITY, ALTHOUGH SOME SUBSITE PREFERENCE HAVE BEEN NOTED,  
 CC E.G. LARGE HYDROPHOBIC RESIDUES IN THE P1 AND P4 POSITIONS, AND  
 CC PRO IN THE P2 POSITION. BEST KNOWN FOR ITS ACTION ON CASEINS,  
 CC ALTHOUGH IT HAS BEEN SHOWN TO HYDROLYSE HEMOGLOBIN AND OXIDIZED  
 CC INSULIN B-CHAIN.  
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE  
 CC SUBTILASE FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC EMBL; X14130; CAA32350.1; -  
 DR PIR; S06997; S06997.  
 DR HSSP; P00782; 2SBT.  
 DR MEROPS; S08.019; -  
 DR InterPro; IPR000209; -

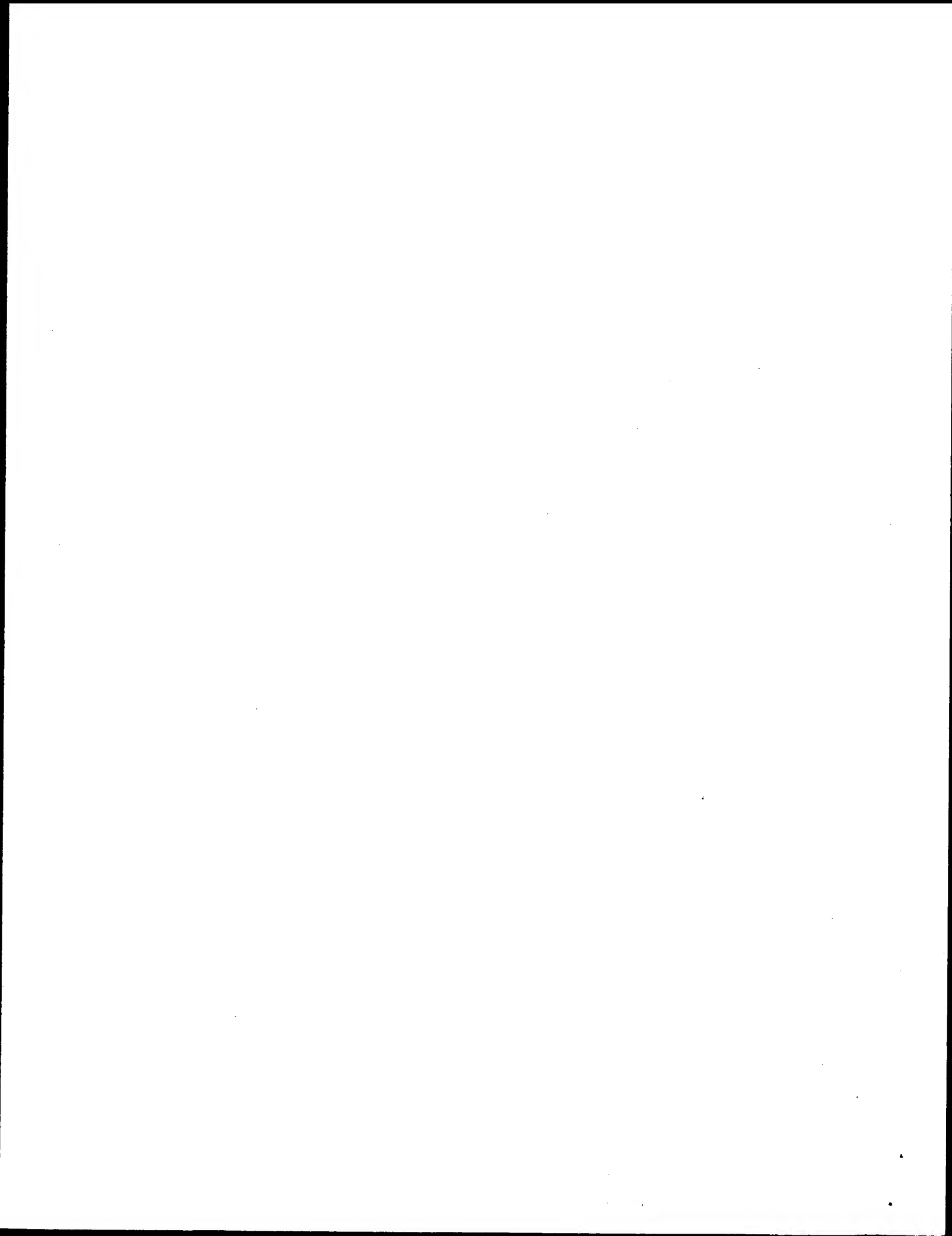






Db 1481 DSEDNTTKNITYEPPKTLAAPTVPSTTEPAKTVTLTANSATGETVQYSADGGKTY 1540  
QY 1509 LNGVGAGANLTAALGVAHSSTHEGKPVGIF---PA--FTSTNVSAALALDNRTSQSISL 1562  
Db 1541 QDVPAAAGVTYTA-----NGTFKFKSTDLYGNESPAVDYVVYVNIKA-----DDPA 1584  
QY 1563 ELKRAEPTVTSNDI--SELTSTLGKHKFDSATTKMLAALKELDDAKPAQOLHILQOHFSAK 1620  
Db 1585 QLQAAKQELTNLIASAKTILSAGRY--DDATTTALAA-----ATQKAQTALDOTNASVD 1636  
QY 1621 DVVGDERYEAVRNLLKLVIRQQAADSHSMELGSASHSTTYNNLSRINNDGIVELLKHIFD 1680  
Db 1637 SLTG-----ANRDLQ-----TAINOLA----- 1653  
QY 1681 AALPASSAKRILGEMMNNDPALKDIIKQLQSTPFSSASYSMELKDGLREQTEK-----A 1733  
Db 1654 AKLPADKKTSL-----LNQLQ-----SVKAALETDLGNQTDSDSTGKTFTA 1693  
QY 1734 ILDGKVGREEVGVLFQDRNNLRVKSVSQSVSKSEGEN--TPALLLGTNSAA 1785  
Db 1694 ALDDLVAQAQAAGTQTDQLOQATLAKVLDVLAFLAKLAEGIKAAATPA-EVGNAKDAA 1746

Search completed: June 5, 2001, 18:23:08  
Job time: 364 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2001, 18:14:34 ; Search time 139.63 Seconds  
(without alignments)  
2117.386 Million cell updates/sec

Title: US-09-596-784-2  
Perfect score: 9448  
Sequence: 1 MELKSLGTEHKAHVHTAAHN.....NPQVASALTDLKKEGLEMK 1838

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1009251 seqs, 160854530 residues

Total number of hits satisfying chosen parameters: 1009251

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending\_Patents\_AA\_Main:\*

1:	/cgn2_6/ptodata/2/paa/PCTUS_COMB.pep.*
2:	/cgn2_6/ptodata/2/paa/US06_COMB.pep.*
3:	/cgn2_6/ptodata/2/paa/US07_COMB.pep.*
4:	/cgn2_6/ptodata/2/paa/US08_COMB.pep.*
5:	/cgn2_6/ptodata/2/paa/US081_COMB.pep.*
6:	/cgn2_6/ptodata/2/paa/US082_COMB.pep.*
7:	/cgn2_6/ptodata/2/paa/US083_COMB.pep.*
8:	/cgn2_6/ptodata/2/paa/US084_COMB.pep.*
9:	/cgn2_6/ptodata/2/paa/US085_COMB.pep.*
10:	/cgn2_6/ptodata/2/paa/US086_COMB.pep.*
11:	/cgn2_6/ptodata/2/paa/US087_COMB.pep.*
12:	/cgn2_6/ptodata/2/paa/US088_COMB.pep.*
13:	/cgn2_6/ptodata/2/paa/US089_COMB.pep.*
14:	/cgn2_6/ptodata/2/paa/US090_COMB.pep.*
15:	/cgn2_6/ptodata/2/paa/US091_COMB.pep.*
16:	/cgn2_6/ptodata/2/paa/US092_COMB.pep.*
17:	/cgn2_6/ptodata/2/paa/US093_COMB.pep.*
18:	/cgn2_6/ptodata/2/paa/US094_COMB.pep.*
19:	/cgn2_6/ptodata/2/paa/US095_COMB.pep.*
20:	/cgn2_6/ptodata/2/paa/US096_COMB.pep.*
21:	/cgn2_6/ptodata/2/paa/US097_COMB.pep.*
22:	/cgn2_6/ptodata/2/paa/US098_COMB.pep.*
23:	/cgn2_6/ptodata/2/paa/US060_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9448	100.0	1838	1 PCT-US98-15426-2	Sequence 2, Appli
2	9448	100.0	1838	15 US-09-120-663-2	Sequence 28, Appl
3	9448	100.0	1838	18 US-09-412-100-28	Sequence 8, Appl
4	9448	100.0	1838	19 US-09-431-614-8	Sequence 2, Appli
5	9448	100.0	1838	19 US-09-596-784-2	Sequence 1, Appli
6	1041	11.0	201	1 PCT-US99-15425-1	Sequence 1, Appli
7	1041	11.0	201	17 US-09-350-852A-1	Sequence 140, App
8	268	2.8	3554	23 US-60-261-974-140	Sequence 4098, Ap
9	257.5	2.7	10203	18 US-09-450-969-4098	Sequence 5564, Ap
10	246	2.6	3351	12 US-08-827-356-5564	

11	246	2.6	3351	20 US-09-611-529-4562	Sequence 4562, Ap
12	242	2.6	2434	23 US-60-242-578-1045	Sequence 1045, Ap
13	242	2.6	2434	23 US-60-253-625-2389	Sequence 2389, Ap
14	242	2.6	2434	23 US-60-257-931-3392	Sequence 3392, Ap
15	242	2.6	2434	23 US-60-269-308-4413	Sequence 4413, Ap
16	240.5	2.5	2086	23 US-60-242-578-948	Sequence 948, App
17	240.5	2.5	2086	23 US-60-253-625-2292	Sequence 2292, Ap
18	240.5	2.5	2086	23 US-60-257-931-3208	Sequence 3208, Ap
19	240.5	2.5	2086	23 US-60-269-308-4231	Sequence 4231, Ap
20	238.5	2.5	2504	17 US-09-328-352-5821	Sequence 5821, Ap
21	236	2.5	3241	19 US-09-558-257-1	Sequence 1, Appli
22	234.5	2.5	1947	12 US-08-827-356-5566	Sequence 5566, Ap
23	234.5	2.5	1947	20 US-09-611-529-4400	Sequence 4400, Ap
24	233	2.5	1996	12 US-08-827-356-3129	Sequence 3129, Ap
25	233	2.5	1996	20 US-09-611-529-6335	Sequence 6335, Ap
26	233	2.5	3696	18 US-09-450-969-5942	Sequence 5942, Ap
27	232	2.5	1981	16 US-09-214-759-38	Sequence 38, Appl
28	226.5	2.4	2025	23 US-60-242-578-990	Sequence 990, App
29	226.5	2.4	2025	23 US-60-253-625-2334	Sequence 2334, Ap
30	226.5	2.4	2025	23 US-60-257-931-3282	Sequence 3282, Ap
31	226.5	2.4	2025	23 US-60-269-308-4306	Sequence 4306, Ap
32	221	2.3	3829	12 US-08-827-356-5567	Sequence 5567, Ap
33	221	2.3	3829	20 US-09-611-529-4591	Sequence 4591, Ap
34	216	2.3	2048	16 US-09-268-347-48	Sequence 48, Appl
35	213.5	2.3	2511	23 US-60-229-518-408	Sequence 408, App
36	213	2.3	1848	23 US-60-173-464-22131	Sequence 22131, A
37	211.5	2.2	1471	23 US-60-215-161-8073	Sequence 8073, Ap
38	211.5	2.2	2680	18 US-09-489-039A-7973	Sequence 7973, Ap
39	210.5	2.2	2423	23 US-60-212-413-181	Sequence 181, App
40	207.5	2.2	2142	19 US-09-540-236-3459	Sequence 3459, Ap
41	207.5	2.2	2142	23 US-60-128-476-4493	Sequence 4493, Ap
42	207.5	2.2	2478	23 US-60-242-578-1042	Sequence 1042, Ap
43	207.5	2.2	2478	23 US-60-253-625-2386	Sequence 2386, Ap
44	207.5	2.2	2478	23 US-60-257-931-3384	Sequence 3384, Ap
45	207.5	2.2	2478	23 US-60-269-308-4405	Sequence 4405, Ap

## ALIGNMENTS

RESULT 1  
PCT-US98-15426-2  
; Sequence 2, Application PC/TUS9815426  
; GENERAL INFORMATION:  
; APPLICANT: Corneil Research Foundation, Inc.  
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR FROM  
; TITLE OF INVENTION: ERWINIA AMYLOVORA, ITS USE, AND ENCODING GENE  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP  
; STREET: P.O. Box 1051, Clinton Square  
; CITY: Rochester  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 14603  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: PC-DOS/MS-DOS  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US98/15426  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/055,105  
; FILING DATE: 06-AUG-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Goldman, Michael L.  
; REGISTRATION NUMBER: 30,727  
; REFERENCE/DOCKET NUMBER: 19603/1662  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (716) 263-1304

TELEFAX: (716) 263-1600  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1838 amino acids  
 TYPE: amino acid  
 STRANDEDNESS:  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 PCT-US98-15426-2

Query Match 100.0%; Score 9448; DB 1; Length 1838;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1838; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MELKSLGTEHKAHVTAHNPVGHGVALOOGSSSSPQNAASLAAGKNGKMPRIHQ 60  
 DB 1 MELKSLGTEHKAHVTAHNPVGHGVALOOGSSSSPQNAASLAAGKNGKMPRIHQ 60

QY 61 STAADGISAHOOOKSFLRGCLGKTKKFSRSPQOGPGTTHSGATLRDLARDGGTQH 120  
 DB 61 STAADGISAHOOOKSFLRGCLGKTKKFSRSPQOGPGTTHSGATLRDLARDGGTQH 120

QY 121 EAAAPDAARLTRSGGKVRNMDMAGRPVKGSGEDKVPYQKRRHOLNPNFQMRQTMLS 180  
 DB 121 EAAAPDAARLTRSGGKVRNMDMAGRPVKGSGEDKVPYQKRRHOLNPNFQMRQTMLS 180

QY 181 KMAHPASANAGDRLOHSPPHIPGSHHEIKEEPPVGSSTKATTAHADRVIEIAQEDDDSEFQ 240  
 DB 181 KMAHPASANAGDRLOHSPPHIPGSHHEIKEEPPVGSSTKATTAHADRVIEIAQEDDDSEFQ 240

QY 241 LHQOARLARENPPPKLGVAATPISARPOKLTAAVESLEGDTTQSPKLPQSMKLG 300  
 DB 241 LHQOARLARENPPPKLGVAATPISARPOKLTAAVESLEGDTTQSPKLPQSMKLG 300

QY 301 GAGVTPLAUTLKGKQLQAPDNPALNTLLKQTLGKDTQHYLAHASSDGSQHLLDNKG 360  
 DB 301 GAGVTPLAUTLKGKQLQAPDNPALNTLLKQTLGKDTQHYLAHASSDGSQHLLDNKG 360

QY 361 HLFDTKSTATSVLHNSHPGETKGLAQAGTGSVDGSGKLSLGSQTSHNKTMLSQ 420  
 DB 361 HLFDTKSTATSVLHNSHPGETKGLAQAGTGSVDGSGKLSLGSQTSHNKTMLSQ 420

QY 421 PGEAHRSLITGTHQHPAGARPQGESIRLHDDKIHLHPELGVWQSAKDQTHSLSQAD 480  
 DB 421 PGEAHRSLITGTHQHPAGARPQGESIRLHDDKIHLHPELGVWQSAKDQTHSLSQAD 480

QY 481 GKLYALKDNRTLQNLSDNKSSEKLVDKIKSYSDQRCQVAILLDTGRRHKSIMPSLDAS 540  
 DB 481 GKLYALKDNRTLQNLSDNKSSEKLVDKIKSYSDQRCQVAILLDTGRRHKSIMPSLDAS 540

QY 541 PESHISLSLHFDADHOGLLHGKSELAQSVASISGRLVWADSEGLFSAATPKQGDGNET 600  
 DB 541 PESHISLSLHFDADHOGLLHGKSELAQSVASISGRLVWADSEGLFSAATPKQGDGNET 600

QY 601 KMKAMPQHALDEHFGHDQISGFFHDDHGLNALVKNFNRQOQHACPLGNDHQFHPGWNLT 660  
 DB 601 KMKAMPQHALDEHFGHDQISGFFHDDHGLNALVKNFNRQOQHACPLGNDHQFHPGWNLT 660

QY 661 DALVIDNQLGLHNTNPEPHEIDMGHLGSLALQEGKLYHFDQLTKGWTGAESDCKQLKKG 720  
 DB 661 DALVIDNQLGLHNTNPEPHEIDMGHLGSLALQEGKLYHFDQLTKGWTGAESDCKQLKKG 720

QY 721 LDGAAYLLKDGKVRNLNINOSTSSIKHGTENVFSLPHVRNKPPEPDALQGLNKDDKAQAM 780  
 DB 721 LDGAAYLLKDGKVRNLNINOSTSSIKHGTENVFSLPHVRNKPPEPDALQGLNKDDKAQAM 780

QY 781 AVIGVKNYIALTEKGDIRSFOIKPGTQOOLRPAQTLREGISGELKDIHVHDKQNYALT 840  
 DB 781 AVIGVKNYIALTEKGDIRSFOIKPGTQOOLRPAQTLREGISGELKDIHVHDKQNYALT 840

QY 841 HGEVEFHOPREAWQNGAESSSWHKLALPQSESKLKLSDMSHEHKPIATFEDGSQHOLKAG 900  
 DB 841 HGEVEFHOPREAWQNGAESSSWHKLALPQSESKLKLSDMSHEHKPIATFEDGSQHOLKAG 900

DB 841 HGEVEFHOPREAWQNGAESSSWHKLALPQSESKLKLSDMSHEHKPIATFEDGSQHOLKAG 900

QY 901 GHAYAAAPERGGLAVCTSGQTVFNRLMQGVKGKVIPIGSGLTVLKLSAQTGNGTGAEGRKV 960  
 DB 901 GHAYAAAPERGGLAVCTSGQTVFNRLMQGVKGKVIPIGSGLTVLKLSAQTGNGTGAEGRKV 960

QY 961 SSKFSERIRAYAFNPMTSTPRPIKNAAYATQHCGWQREGKLPYEMOGALIKOLDAHNVR 1020  
 DB 961 SSKFSERIRAYAFNPMTSTPRPIKNAAYATQHCGWQREGKLPYEMOGALIKOLDAHNVR 1020

QY 1021 HNAPODQLOSKLETLDLGEHGAELLNDMKFRDELEQSATRSVTVLGOHGVKLKNGEIN 1080  
 DB 1021 HNAPODQLOSKLETLDLGEHGAELLNDMKFRDELEQSATRSVTVLGOHGVKLKNGEIN 1080

QY 1081 SEFKPSGKALVQSFNVNRSQDLKSLQOAVHATPPSAESKLSQMLGHFVSAGVDMSHQ 1140  
 DB 1081 SEFKPSGKALVQSFNVNRSQDLKSLQOAVHATPPSAESKLSQMLGHFVSAGVDMSHQ 1140

QY 1141 KGEIPLGRQDPNDKTALTAKSRLILDTVTIGELHELADKAKLVSDHKPDADOIKOLRQOF 1200  
 DB 1141 KGEIPLGRQDPNDKTALTAKSRLILDTVTIGELHELADKAKLVSDHKPDADOIKOLRQOF 1200

QY 1201 DTLREKRYESNPVKKHYTDMGFTHNKALEANYDAVKAFINAEKKEHGVNLTTRTVLESOG 1260  
 DB 1201 DTLREKRYESNPVKKHYTDMGFTHNKALEANYDAVKAFINAEKKEHGVNLTTRTVLESOG 1260

QY 1261 SAEAKKLKNTLLSDGESMSFSRSYGGGVSTVFVPTLSKKVPVPIPGAGITLDRAIN 1320  
 DB 1261 SAEAKKLKNTLLSDGESMSFSRSYGGGVSTVFVPTLSKKVPVPIPGAGITLDRAIN 1320

QY 1321 LFSRSTSGGLNVSGFRDGGVSGNIMVATGHDVMPYMTGKTSAGNASDWLSAKHKISPDL 1380  
 DB 1321 LFSRSTSGGLNVSGFRDGGVSGNIMVATGHDVMPYMTGKTSAGNASDWLSAKHKISPDL 1380

QY 1381 RIGAAVSGTLQGLTQNSLKEFLTEDELPGFTHGLTHGTLTPAELLQKGLIEHOMKQSKLT 1440  
 DB 1381 RIGAAVSGTLQGLTQNSLKEFLTEDELPGFTHGLTHGTLTPAELLQKGLIEHOMKQSKLT 1440

QY 1441 FSVDTSANLDRAGINLNEGSKPNGVTARYSAGLSASANLAAGSRERSTTSQGFSGTTS 1500  
 DB 1441 FSVDTSANLDRAGINLNEGSKPNGVTARYSAGLSASANLAAGSRERSTTSQGFSGTTS 1500

QY 1501 ASNNRPTFLNGVGAGANLTAALGVASHSTHEGKPVGIFPAFTSTNVSAALADNRTSQSI 1560  
 DB 1501 ASNNRPTFLNGVGAGANLTAALGVASHSTHEGKPVGIFPAFTSTNVSAALADNRTSQSI 1560

QY 1561 SLELKRAEPTVNDISELTSTLGHKFKDSATTKMLAALKELDDAKPAEQLHILQOHFSK 1620  
 DB 1561 SLELKRAEPTVNDISELTSTLGHKFKDSATTKMLAALKELDDAKPAEQLHILQOHFSK 1620

QY 1621 DVVGDERYEAARNLKLVIROQAADSHSMELGSHSTTYNNLSRINNDGIVELLKHFD 1680  
 DB 1621 DVVGDERYEAARNLKLVIROQAADSHSMELGSHSTTYNNLSRINNDGIVELLKHFD 1680

QY 1681 AALPASSAKRLGEMMNDPALKDIILQOSTPFSSASVSMELKDGLREOTEKAILDGKVG 1740  
 DB 1681 AALPASSAKRLGEMMNDPALKDIILQOSTPFSSASVSMELKDGLREOTEKAILDGKVG 1740

QY 1741 REEVGVLFQDRNNLRVKSVSQSVKSEGFNTPALLLGTSSAAMSMEINIGTINPKFY 1800  
 DB 1741 REEVGVLFQDRNNLRVKSVSQSVKSEGFNTPALLLGTSSAAMSMEINIGTINPKFY 1800

QY 1801 QDONTPRRFTLEGGIAQANPOVASALTDLKKEGLEMK 1838  
 DB 1801 QDONTPRRFTLEGGIAQANPOVASALTDLKKEGLEMK 1838

RESULT 2  
 US-09-120-663-2  
 ; Sequence 2, Application US/09120663  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bogdanove, Adam J.  
 ; APPLICANT: Kim, Jihyun Francis

```

; APPLICANT: Wei, Zhong-Min
; APPLICANT: Beer, Steven V.
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR FROM
; TITLE OF INVENTION: ERWINIA AMYLOVORA, ITS USE, AND ENCODING GENE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: P.O. Box 1051, Clinton Square
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/120,663
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/055,105
; FILING DATE: 06-AUG-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/1661
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1838 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
US-09-120-663-2

```

```

Query Match      100.0%; Score 9448; DB 15; Length 1838;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1838; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MELKSLGTEHKAHVHTAAHNPVGHGVALQOQSSSSPQNAASLAAGKNGKMPRIHQ 60
Db 1 MELKSLGTEHKAHVHTAAHNPVGHGVALQOQSSSSPQNAASLAAGKNGKMPRIHQ 60

Qy 61 STAADGISAHQAQKSFSLRCLGCTKKFSRPAQOGPGTTHSKGATLRDLLARDGETQH 120
Db 61 STAADGISAHQAQKSFSLRCLGCTKKFSRPAQOGPGTTHSKGATLRDLLARDGETQH 120

Qy 121 EAAAPDAARLTRSGVVKRRNMDMAGRPWKGGSGEDKVPQOQRHQLNFGQMRQTMLS 180
Db 121 EAAAPDAARLTRSGVVKRRNMDMAGRPWKGGSGEDKVPQOQRHQLNFGQMRQTMLS 180

Qy 181 KMAHPASANACDRLOHSPPHIPGSHHETKEEPPVGSSTKATTAHADRVIAQEDDDSEFQ 240
Db 181 KMAHPASANACDRLOHSPPHIPGSHHETKEEPPVGSSTKATTAHADRVIAQEDDDSEFQ 240

Qy 241 LHQRLARERENPPQPKLGVA TPISARFQPKLTAVAESVLEGTDTTQSPKLPQSMKGS 300
Db 241 LHQRLARERENPPQPKLGVA TPISARFQPKLTAVAESVLEGTDTTQSPKLPQSMKGS 300

Qy 301 GAGVTPLATVLDKGLQALAPONPPALNTLLKQTLGKDTQHYLAHHASDGSQHLDDNKG 360
Db 301 GAGVTPLATVLDKGLQALAPONPPALNTLLKQTLGKDTQHYLAHHASDGSQHLDDNKG 360

Qy 361 HLFDIKSTATSYSVLHNSHPGEIKGLAQAGTGSVSDGKSGKISLGSGTQSHNKTMLSQ 420
Db 361 HLFDIKSTATSYSVLHNSHPGEIKGLAQAGTGSVSDGKSGKISLGSGTQSHNKTMLSQ 420

```

```

Qy 421 PGEAHRSLTGTWHPAGAAARPGQESIRLHDDKIHILHPELVWQADKDTHSQLSROAD 480
Db 421 PGEAHRSLTGTWHPAGAAARPGQESIRLHDDKIHILHPELVWQADKDTHSQLSROAD 480
Qy 481 GKLYALKDNRTNLQNLSDNKSSEKLVKIKSYSDQGOVAILTDTTPGRHKMSIMPSLDAS 540
Db 481 GKLYALKDNRTNLQNLSDNKSSEKLVKIKSYSDQGOVAILTDTTPGRHKMSIMPSLDAS 540
Qy 541 PESHISLSLHFAHOGHLLHGKSELEAQSVAISHGRVLVADSEKGLFSAAPKOGDGNEL 600
Db 541 PESHISLSLHFAHOGHLLHGKSELEAQSVAISHGRVLVADSEKGLFSAAPKOGDGNEL 600
Qy 601 KMKAMPQHALDEHFHGHQISGFFHDDHGNALVKNFNQOQACPLGNDHQHFGPNLNT 660
Db 601 KMKAMPQHALDEHFHGHQISGFFHDDHGNALVKNFNQOQACPLGNDHQHFGPNLNT 660
Qy 661 DALVIDNQLGLHHTNPPEHIELDMGLSLALQEGKIHYPDQLTKGWTGAESCKQLKKG 720
Db 661 DALVIDNQLGLHHTNPPEHIELDMGLSLALQEGKIHYPDQLTKGWTGAESCKQLKKG 720
Qy 721 LDGAAYLLKDGVEVKRLNINQSTSSIKHGTENVSFLPHVRNKPPEGDALQGLNKDDKAAQ 780
Db 721 LDGAAYLLKDGVEVKRLNINQSTSSIKHGTENVSFLPHVRNKPPEGDALQGLNKDDKAAQ 780
Qy 781 AVIGVKNYLALTEKGDIRSFOIKPGTOOLERPACTLSREGISGELKDIHVDHKQNLVLT 840
Db 781 AVIGVKNYLALTEKGDIRSFOIKPGTOOLERPACTLSREGISGELKDIHVDHKQNLVLT 840
Qy 841 HEGEVFHPREAWONGAESSESHKLLALPQSESKLKLSDMSHEHKPIATFEDGSOHLKAG 900
Db 841 HEGEVFHPREAWONGAESSESHKLLALPQSESKLKLSDMSHEHKPIATFEDGSOHLKAG 900
Qy 901 GWHAYAAPERGPLAVGTSGSQTVEFNRLMQGVKGVIPGSGLTVKLSAQTGGMGTGAEGRV 960
Db 901 GWHAYAAPERGPLAVGTSGSQTVEFNRLMQGVKGVIPGSGLTVKLSAQTGGMGTGAEGRV 960
Qy 961 SSKFSEIRAYAPNPTMTSTPRPIKNAAYATOHGQREGCLKPLVEMQCALIKOLDAAHVR 1020
Db 961 SSKFSEIRAYAPNPTMTSTPRPIKNAAYATOHGQREGCLKPLVEMQCALIKOLDAAHVR 1020
Qy 1021 HNAPQDPLQSKLETLDLGEHGAELLNDMKRDELEQATRSVTVLGQHGVLKNGEIN 1080
Db 1021 HNAPQDPLQSKLETLDLGEHGAELLNDMKRDELEQATRSVTVLGQHGVLKNGEIN 1080
Qy 1081 SEFKPSGKALVQSFNVNRSGDLSKSLQQAQVHATPPSAESKLSMLCHFFVSAGVDMSHQ 1140
Db 1081 SEFKPSGKALVQSFNVNRSGDLSKSLQQAQVHATPPSAESKLSMLCHFFVSAGVDMSHQ 1140
Qy 1141 KGEIPLGRQRPNDKTALTAKSRLILDVTYIGELHELADKALVSDHKPDADQIKOLROOF 1200
Db 1141 KGEIPLGRQRPNDKTALTAKSRLILDVTYIGELHELADKALVSDHKPDADQIKOLROOF 1200
Qy 1201 DTLREKRYESNPVKHYTDMGFTTHNKALEANYDAVKAFINAFKKEHGVNLTTRTVLESOG 1260
Db 1201 DTLREKRYESNPVKHYTDMGFTTHNKALEANYDAVKAFINAFKKEHGVNLTTRTVLESOG 1260
Qy 1261 SAEIAKKLUNTLNLLSDGESMSFSRSYGGGVSTVFPVTLKKVPVPIPCAGITLDRAVN 1320
Db 1261 SAEIAKKLUNTLNLLSDGESMSFSRSYGGGVSTVFPVTLKKVPVPIPCAGITLDRAVN 1320
Qy 1321 LSFSTSGGLNVSGRQDGVSGNIMVATGHDVMPYMTGKTSAGNASDWLSAKHKISPD 1380
Db 1321 LSFSTSGGLNVSGRQDGVSGNIMVATGHDVMPYMTGKTSAGNASDWLSAKHKISPD 1380
Qy 1381 RIGAAVSGTLOQTLONSLKFILTEDELPGFTHGLTHGLTTPAELLQKGIHEHOMKQSKLT 1440
Db 1381 RIGAAVSGTLOQTLONSLKFILTEDELPGFTHGLTHGLTTPAELLQKGIHEHOMKQSKLT 1440
Qy 1441 FSVDTSANILDRAGINLNDGSKPNCVTVARSAGLSASANLAAGSRERSTTSGQFGSTTS 1500
Db 1441 FSVDTSANILDRAGINLNDGSKPNCVTVARSAGLSASANLAAGSRERSTTSGQFGSTTS 1500
Qy 1501 ASNNRPTFLNGVGAGANLTAALGVAHSSHTHEGKPVGIFPAFTSTNVSAALADNRSTOSI 1560

```

Db 1501 ASNNRPTFLNGVAGANLTAALGAHVSSTHEGKPVGIFPAFTSNVSAALALDNRTSQSI 1560  
QY 1561 SLELKRAEPTVNDISELSTGLKHKFDSATTWMLAALKELDPAEQHILLOQHFSAK 1620  
Db 1561 SLELKRAEPTVNDISELSTGLKHKFDSATTWMLAALKELDPAEQHILLOQHFSAK 1620  
QY 1621 DVVGDERYEAVRNKLKLVIRQQAADSHSMELGSAHSHTTYNNLSRINNDGIVELLHKHFD 1680  
Db 1621 DVVGDERYEAVRNKLKLVIRQQAADSHSMELGSAHSHTTYNNLSRINNDGIVELLHKHFD 1680  
QY 1681 AALPASSAKRIGEMMNDPALKDIIKOLQSTPSSASVSMELKDLGLEQTEKAILDGKVG 1740  
Db 1681 AALPASSAKRIGEMMNDPALKDIIKOLQSTPSSASVSMELKDLGLEQTEKAILDGKVG 1740  
QY 1741 REEVGLVFDNRNLRKVSQSVSKSGFNTPALLGTSNSAAMSMERNIGTINPKYG 1800  
Db 1741 REEVGLVFDNRNLRKVSQSVSKSGFNTPALLGTSNSAAMSMERNIGTINPKYG 1800  
QY 1801 QDONTPTRETTLEGGIAQANPOVASALTDLKKEGLEMS 1838  
Db 1801 QDONTPTRETTLEGGIAQANPOVASALTDLKKEGLEMS 1838

RESULT 3  
US-09-412-100-28  
; Sequence 28, Application US/09412100  
; GENERAL INFORMATION:  
; APPLICANT: Wei, Zhong-Min  
; APPLICANT: Fan, Hao  
; APPLICANT: Niggemeyer, Jennifer L.  
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR FRAGMENTS WHICH ARE  
; FILE OF INVENTION: ACTIVE BUT DO NOT ELICIT A HYPERSENSITIVE RESPONSE  
; FILE REFERENCE: 21829/31 (EBC-002)  
; CURRENT APPLICATION NUMBER: US/09/412,100  
; CURRENT FILING DATE: 1999-10-04  
; EARLIER APPLICATION NUMBER: 60/103,050  
; EARLIER FILING DATE: 1998-10-05  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: PatentIn ver. 2.0  
; SEQ ID NO 28  
; LENGTH: 1838  
; TYPE: PRT  
; ORGANISM: Erwinia amylovora  
US-09-412-100-28

Query Match 100.0%; Score 9448; DB 18; Length 1838;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1838; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MELKSLGTEHKAHVHTAAHNPVGHVALQOGSSSSSPONAAAASLAAGKNGKMPRIHQ 60  
Db 1 MELKSLGTEHKAHVHTAAHNPVGHVALQOGSSSSSPONAAAASLAAGKNGKMPRIHQ 60  
QY 61 STAADGISAHQKKSFLRGCLGCTKFSRSPQOGCTTHSKGATLRDLLARDGGETQH 120  
Db 61 STAADGISAHQKKSFLRGCLGCTKFSRSPQOGCTTHSKGATLRDLLARDGGETQH 120  
QY 121 EAAAPDAARLTRSGVYRRNDDMAGRPVKGSGEDKVPYQQRHQLNFGQMRQTMLS 180  
Db 121 EAAAPDAARLTRSGVYRRNDDMAGRPVKGSGEDKVPYQQRHQLNFGQMRQTMLS 180  
QY 181 KMAHPASANAGDLRHSPPHPPSHHIEKEEVPGSTSKATTAHADRVETIAQEDDDSEFQQ 240  
Db 181 KMAHPASANAGDLRHSPPHPPSHHIEKEEVPGSTSKATTAHADRVETIAQEDDDSEFQQ 240  
QY 241 LHOQRLARENPQPPLKGVATPISARFQPKLTAVAESVLEGTDTTQSPCLKPQSMCLKGS 300  
Db 241 LHOQRLARENPQPPLKGVATPISARFQPKLTAVAESVLEGTDTTQSPCLKPQSMCLKGS 300  
QY 301 GAGVTPPLAVTLDKGLQLAPDNPALNTLLKQTLGKDTQHYLAHASSDGSQHLLLDNKG 360  
Db 301 GAGVTPPLAVTLDKGLQLAPDNPALNTLLKQTLGKDTQHYLAHASSDGSQHLLLDNKG 360

Db 301 GAGVTPPLAVTLDKGLQLAPDNPALNTLLKQTLGKDTQHYLAHASSDGSQHLLLDNKG 360  
QY 361 HLFDTIKSTATSYSLHNSHPGEIKGLAQAGTGSVSDGSKGISLGSQTSQHNKTMLSQ 420  
Db 361 HLFDTIKSTATSYSLHNSHPGEIKGLAQAGTGSVSDGSKGISLGSQTSQHNKTMLSQ 420  
QY 421 PGEAHRSLTGTIOWHPAGAARPOGESIRLHDDKIHILHPPELGVQWQADKTHSOLRQAD 480  
Db 421 PGEAHRSLTGTIOWHPAGAARPOGESIRLHDDKIHILHPPELGVQWQADKTHSOLRQAD 480  
QY 481 GKLYALKDNRTLQNLSDNKSSEKLVDKIKSYSDVORGQVAILTDTTPGRHKMSIMPSLDAS 540  
Db 481 GKLYALKDNRTLQNLSDNKSSEKLVDKIKSYSDVORGQVAILTDTTPGRHKMSIMPSLDAS 540  
QY 541 PSHISLSLHPADAHQGLLHGKSELEAQSVAISHGLRVVADSEGLFSAAIKPGQDGNEL 600  
Db 541 PSHISLSLHPADAHQGLLHGKSELEAQSVAISHGLRVVADSEGLFSAAIKPGQDGNEL 600  
QY 601 KMKAMPOHALDEHFGHDHIOISGFFHDDHQLNALYKNNFRQOHACPLGNDHOFHPGWNLT 660  
Db 601 KMKAMPOHALDEHFGHDHIOISGFFHDDHQLNALYKNNFRQOHACPLGNDHOFHPGWNLT 660  
QY 661 DALVIDNQLGLHHTNPEPHEITLDMGHLGSLALQEGKHYFDOLTGKWTGAESDCQKQK 720  
Db 661 DALVIDNQLGLHHTNPEPHEITLDMGHLGSLALQEGKHYFDOLTGKWTGAESDCQKQK 720  
QY 721 LDGAAYLLKDGCEVKRLINQSTSSIKHGTENVFSLPHVNRKPEPGDALQGLNKKDKAQAM 780  
Db 721 LDGAAYLLKDGCEVKRLINQSTSSIKHGTENVFSLPHVNRKPEPGDALQGLNKKDKAQAM 780  
QY 781 AVIGVKNYKALTEKGDIRSFOIKPTQOLERPACTLSREGISGELKDIHVHDKQNLALT 840  
Db 781 AVIGVKNYKALTEKGDIRSFOIKPTQOLERPACTLSREGISGELKDIHVHDKQNLALT 840  
QY 841 HEGEVHPQREAWONGAESSWHKLALPQSESKLSLSDMSHEHKPIATFEDGSOHQKAG 900  
Db 841 HEGEVHPQREAWONGAESSWHKLALPQSESKLSLSDMSHEHKPIATFEDGSOHQKAG 900  
QY 901 GWHAYAAPERGPLAVGTSGSOTVFENRLMGVKGKVIKSGSLTVKLSAOTGGTGAERKV 960  
Db 901 GWHAYAAPERGPLAVGTSGSOTVFENRLMGVKGKVIKSGSLTVKLSAOTGGTGAERKV 960  
QY 961 SSKFESIRIAYAFNPMTSTPRPKNAAYATQHGQWQREGLKPLXYEMOGALIKQLDAHNR 1020  
Db 961 SSKFESIRIAYAFNPMTSTPRPKNAAYATQHGQWQREGLKPLXYEMOGALIKQLDAHNR 1020  
QY 1021 HNAPODLOSKELETDLGEGHAELLNDMKRFRDELEQSATRSVTVLGOHQGVLSNGEIN 1080  
Db 1021 HNAPODLOSKELETDLGEGHAELLNDMKRFRDELEQSATRSVTVLGOHQGVLSNGEIN 1080  
QY 1081 SEFKPSPGKALVQSFVNNRSGQDLSKSLQQAQVHATPPSAESKLSQMLGCHFVSAGVDMSHQ 1140  
Db 1081 SEFKPSPGKALVQSFVNNRSGQDLSKSLQQAQVHATPPSAESKLSQMLGCHFVSAGVDMSHQ 1140  
QY 1141 KGEIPLGRDRDPNDKALTAKSRLLDVTITGELHELADKAKLVSDHDPDADQIKOLRQOF 1200  
Db 1141 KGEIPLGRDRDPNDKALTAKSRLLDVTITGELHELADKAKLVSDHDPDADQIKOLRQOF 1200  
QY 1201 DTLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKKEHHGVNLTTRTVLESOG 1260  
Db 1201 DTLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKKEHHGVNLTTRTVLESOG 1260  
QY 1261 SAEALAKKNTLLSDSGESMSFSRSYGGGVSTFVPTLSKKVPVPVPIPGAGITLDRAYN 1320  
Db 1261 SAEALAKKNTLLSDSGESMSFSRSYGGGVSTFVPTLSKKVPVPVPIPGAGITLDRAYN 1320  
QY 1321 LSFSTSGGLNVSGFRDGGVSGNIMVATGHDVMPYMGKTTSAGNASDWLSAKHKISPDL 1380  
Db 1321 LSFSTSGGLNVSGFRDGGVSGNIMVATGHDVMPYMGKTTSAGNASDWLSAKHKISPDL 1380  
QY 1381 RIGAAVSGTLOQTLONSLKFKLTDELPGF IHGLTHTGLTTPAEELLQKIEHQHKKQSKLT 1440  
Db 1381 RIGAAVSGTLOQTLONSLKFKLTDELPGF IHGLTHTGLTTPAEELLQKIEHQHKKQSKLT 1440

QY 1441 FSDVTSANLDLDRAGINLNEDGSKPNCVTVARVSAGLSASANLAAGSRERSTTSQOFGSTTS 1500  
DB 1441 FSDVTSANLDLDRAGINLNEDGSKPNCVTVARVSAGLSASANLAAGSRERSTTSQOFGSTTS 1500  
QY 1501 ASNNRPTFLNGVAGANLTAALGVVAHSSTHEGKPGVIGFPAFTSNVSAALALDNRTSQSI 1560  
DB 1501 ASNNRPTFLNGVAGANLTAALGVVAHSSTHEGKPGVIGFPAFTSNVSAALALDNRTSQSI 1560  
QY 1561 SLELKRAEPTVNDISELSTLGLKHKFDSATTKMLAALKELDDAKPAEQHLLHQHFSAK 1620  
DB 1561 SLELKRAEPTVNDISELSTLGLKHKFDSATTKMLAALKELDDAKPAEQHLLHQHFSAK 1620  
QY 1621 DVYGERYEAVERNKLKLVTRQQAADSHSMELGSAHSSTYNNLSRINNDGIVELLHKHFD 1680  
DB 1621 DVYGERYEAVERNKLKLVTRQQAADSHSMELGSAHSSTYNNLSRINNDGIVELLHKHFD 1680  
QY 1681 AALPASSAKRLGEMMNDPALKDITIKOLOSTPFSSASVSMELKDGLREQTEKATLDCKVG 1740  
DB 1681 AALPASSAKRLGEMMNDPALKDITIKOLOSTPFSSASVSMELKDGLREQTEKATLDCKVG 1740  
QY 1741 REEVGYLFODRNNLRKVSVSQSVSKSEGFNTPALLLGTSNAAAMSWERNIGTINPKYG 1800  
DB 1741 REEVGYLFODRNNLRKVSVSQSVSKSEGFNTPALLLGTSNAAAMSWERNIGTINPKYG 1800  
QY 1801 QDONTPRRTLEGGIAQANPOVASALTDLKKEGLEMS 1838  
DB 1801 QDONTPRRTLEGGIAQANPOVASALTDLKKEGLEMS 1838

## RESULT 4

US-09-431-614-8

; Sequence 8, Application US/09431614

; GENERAL INFORMATION:

; APPLICANT: Wei, Zhong-Min

; APPLICANT: Schading, Richard L.

; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR-INDUCED STRESS

; TITLE OF INVENTION: RESISTANCE

; FILE REFERENCE: 21829/41 (EBC-003)

; CURRENT APPLICATION NUMBER: US/09/431,614

; EARLIER FILING DATE: 1999-11-02

; EARLIER APPLICATION NUMBER: 60/107,243

; EARLIER FILING DATE: 1998-11-05

; NUMBER OF SEQ ID NOS: 18

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 8

; LENGTH: 1838

; TYPE: PRT

; ORGANISM: Erwinia amylovora

US-09-431-614-8

Query Match 100.0%; Score 9448; DB 18; Length 1838;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1838; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MELKSLGTEHKAHVTAHNPVGHVVALQOQSSSSPONAASLAAGCKNKGKMPRIHQ 60  
DB 1 MELKSLGTEHKAHVTAHNPVGHVVALQOQSSSSPONAASLAAGCKNKGKMPRIHQ 60  
QY 61 STAADGISAHOQKSFSLRGCLCTKFSRPAQOGPGTTHSKGATLRDLARDDETQH 120  
DB 61 STAADGISAHOQKSFSLRGCLCTKFSRPAQOGPGTTHSKGATLRDLARDDETQH 120  
QY 121 EAAAPDAARLTRSGGVKRRNMDMAGRPVKGSGEDKVPVTOQKRHLQNNFGQMRTMLS 180  
DB 121 EAAAPDAARLTRSGGVKRRNMDMAGRPVKGSGEDKVPVTOQKRHLQNNFGQMRTMLS 180  
QY 181 KMAHPASAGDRLOHSPPHIPGSHHEIKEEPVSGTSKATTAHADRVETAQEDDSEFOQ 240  
DB 181 KMAHPASAGDRLOHSPPHIPGSHHEIKEEPVSGTSKATTAHADRVETAQEDDSEFOQ 240  
QY 241 LHOORLARERENPPPKLVGATPISARFQPKLTAVAESVLEGTDTTQSPKQSMKGS 300

DB 241 LHOORLARERENPPPKLVGATPISARFQPKLTAVAESVLEGTDTTQSPKQSMKGS 300  
QY 301 GAGVTPLATVLDKGLQAPDNPPALNTLLKQTLGKDTQHYLAHASSDGSQHLLDNKG 360  
DB 301 GAGVTPLATVLDKGLQAPDNPPALNTLLKQTLGKDTQHYLAHASSDGSQHLLDNKG 360  
QY 361 HLFDLKSTATSYSVLHNSHPGEIKKLAQAGTGSVDGKSGKISLGSGTQSHNKTMLSQ 420  
DB 361 HLFDLKSTATSYSVLHNSHPGEIKKLAQAGTGSVDGKSGKISLGSGTQSHNKTMLSQ 420  
QY 421 PGEAHRSLLTGTHWHPAGAARPOGESIRLHDDKIHILHPELGVMSADKDTHSQLSROAD 480  
DB 421 PGEAHRSLLTGTHWHPAGAARPOGESIRLHDDKIHILHPELGVMSADKDTHSQLSROAD 480  
QY 481 GKLYALKDNRTLONLSDNKSSEKLVYDKIKSYSDQRCQVAILTDTPCRHKMSIMPSLDAS 540  
DB 481 GKLYALKDNRTLONLSDNKSSEKLVYDKIKSYSDQRCQVAILTDTPCRHKMSIMPSLDAS 540  
QY 541 PESHTLSLHFADAHQGLLRGKSELEAQSVAISHGRLLVADSEGLFSAAPKOGDGNEL 600  
DB 541 PESHTLSLHFADAHQGLLRGKSELEAQSVAISHGRLLVADSEGLFSAAPKOGDGNEL 600  
QY 601 KMKAMPOHALDEHFGHDHQSISGFFHDDHGOALNVLNFRQOQACPLGNDHQPHPGNLT 660  
DB 601 KMKAMPOHALDEHFGHDHQSISGFFHDDHGOALNVLNFRQOQACPLGNDHQPHPGNLT 660  
QY 661 DALVIDNQLGHLHTNPEPHEILDMGHLSLALQEGKLHYFDQTLKGTGAECDCKQLKKG 720  
DB 661 DALVIDNQLGHLHTNPEPHEILDMGHLSLALQEGKLHYFDQTLKGTGAECDCKQLKKG 720  
QY 721 LDGAAYLLKXGDEVKRLNINOSTSSIKHGTENFVSLPHVRNKPPEPDALOGLNKDDKAQAM 780  
DB 721 LDGAAYLLKXGDEVKRLNINOSTSSIKHGTENFVSLPHVRNKPPEPDALOGLNKDDKAQAM 780  
QY 781 AVIGVKNYLLATEKGDIRSFOIKPGTQOOLRPAQTLSREGISELKDIDVDHKNLYALT 840  
DB 781 AVIGVKNYLLATEKGDIRSFOIKPGTQOOLRPAQTLSREGISELKDIDVDHKNLYALT 840  
QY 841 HEGEVPHOPREAWONGAESSSWHKLLALPQSESKLSLDMSHHEKPIATFEDGSOHLKAG 900  
DB 841 HEGEVPHOPREAWONGAESSSWHKLLALPQSESKLSLDMSHHEKPIATFEDGSOHLKAG 900  
QY 901 GWHYAAAPERGPLAVGTSGSTVFNRLMOGVKGVIPGSGLTVTKLSAQGTGWTGAERKV 960  
DB 901 GWHYAAAPERGPLAVGTSGSTVFNRLMOGVKGVIPGSGLTVTKLSAQGTGWTGAERKV 960  
QY 961 SSKFSERIAYAFNPTMSTPRPIKNAAYATOHGQREGKLPLEYEMOGALLKOLDAHNVR 1020  
DB 961 SSKFSERIAYAFNPTMSTPRPIKNAAYATOHGQREGKLPLEYEMOGALLKOLDAHNVR 1020  
QY 1021 HNAPODLQSKLETLDLGEHCAELLNDMKFRDELEQSATRSVTVLGHOGLVKSNGEIN 1080  
DB 1021 HNAPODLQSKLETLDLGEHCAELLNDMKFRDELEQSATRSVTVLGHOGLVKSNGEIN 1080  
QY 1081 SEFKPSGKALVGSFNVNRSQDLKSLQQAQVHATPPSAESKLSQSMGLGHFVSAGVDMHSQ 1140  
DB 1081 SEFKPSGKALVGSFNVNRSQDLKSLQQAQVHATPPSAESKLSQSMGLGHFVSAGVDMHSQ 1140  
QY 1141 KGEIPIGRQDPNDKTALTYSRLILDTVTIGELHELADKALVSDHKFPDADQIKOLROOF 1200  
DB 1141 KGEIPIGRQDPNDKTALTYSRLILDTVTIGELHELADKALVSDHKFPDADQIKOLROOF 1200  
QY 1201 DTLREKRYESNPVKHYTDMGTTHNKALNDAVKAFINAFKKEHGHVNLTRTRVLESQ 1260  
DB 1201 DTLREKRYESNPVKHYTDMGTTHNKALNDAVKAFINAFKKEHGHVNLTRTRVLESQ 1260  
QY 1261 SAEKAKKLTLLSLDSGESMSFSRSYGGVSTVFVPTLSKKVPVPIPGAGITLDRAYN 1320  
DB 1261 SAEKAKKLTLLSLDSGESMSFSRSYGGVSTVFVPTLSKKVPVPIPGAGITLDRAYN 1320  
QY 1321 LSPFSRTSGGLUNVSGRGGVSGNTMWTGHDVMPYMTGKKTTSAGNASDWLSAKHKISDPL 1380  
DB 1321 LSPFSRTSGGLUNVSGRGGVSGNTMWTGHDVMPYMTGKKTTSAGNASDWLSAKHKISDPL 1380



Db 1321 LSFSTSGGLNVFGRDGGVGNIMVATGHDVMPYMTGKKTSGNASDWLSAKHKISPDL 1380  
QY 1381 RIGAVSGTLOQTLONSLKFILTEDELPGFIHGLTHGTLTPAELLQKGIHQHOMKQGSKLT 1440  
Db 1381 RIGAVSGTLOQTLONSLKFILTEDELPGFIHGLTHGTLTPAELLQKGIHQHOMKQGSKLT 1440  
QY 1441 FSVDTSANLDIRAGINLNEDESKPGNVTVARVSGLSASANLAAGSRERSTTSQGFSGTTS 1500  
Db 1441 FSVDTSANLDIRAGINLNEDESKPGNVTVARVSGLSASANLAAGSRERSTTSQGFSGTTS 1500  
QY 1501 ASNNRPTFLNGVAGANLTAALGVAHSHHEGKPGVCIIPAFSTNVSAALDNRTSQSI 1560  
Db 1501 ASNNRPTFLNGVAGANLTAALGVAHSHHEGKPGVCIIPAFSTNVSAALDNRTSQSI 1560  
QY 1561 SLELKRAEPTVNDISELTSTLTKHFKDSATTKMLAALKELDDAKPAEOLHILQOHFSK 1620  
Db 1561 SLELKRAEPTVNDISELTSTLTKHFKDSATTKMLAALKELDDAKPAEOLHILQOHFSK 1620  
QY 1621 DVVGDERYEAVRNLLKLVIRQAADSHSMELGSAHSHSTYNNLSRINNDGIVELLHKHFD 1680  
Db 1621 DVVGDERYEAVRNLLKLVIRQAADSHSMELGSAHSHSTYNNLSRINNDGIVELLHKHFD 1680  
QY 1681 AALPASSAKRLGEMNNDPALDKIIKQLQSTPFSSASVSMELKDGIREOTEKAILDKGVG 1740  
Db 1681 AALPASSAKRLGEMNNDPALDKIIKQLQSTPFSSASVSMELKDGIREOTEKAILDKGVG 1740  
QY 1741 REEVGVLFQDRNNLRVKSVSQSVKSEGFNTPALLLGTSNSAAMSMERNIGTINFKYG 1800  
Db 1741 REEVGVLFQDRNNLRVKSVSQSVKSEGFNTPALLLGTSNSAAMSMERNIGTINFKYG 1800  
QY 1801 QDQNTPRFTLEGGAQANPOVASALTDLKEGLEMK 1838  
Db 1801 QDQNTPRFTLEGGAQANPOVASALTDLKEGLEMK 1838

## RESULT 5

US-09-596-784-2  
: Sequence 2, Application US/09596784  
: GENERAL INFORMATION:  
: APPLICANT: Bogdanove, Adam J.  
: APPLICANT: Kim, Jihyun Francis  
: APPLICANT: Wei, Zhong-Min  
: APPLICANT: Beer, Steven V.  
: TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR FROM  
: NUMBER OF SEQUENCES: 5  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP  
: STREET: P.O. Box 1051, Clinton Square  
: CITY: Rochester  
: STATE: New York  
: COUNTRY: U.S.A.  
: ZIP: 14603  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: PatentIn Release #1.0, Version #1.30  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/09/596,784  
: FILING DATE:  
: CLASSIFICATION:  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: 09/120,663  
: FILING DATE:  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Goldman, Michael L.  
: REGISTRATION NUMBER: 30,727  
: REFERENCE/DOCKET NUMBER: 19603/1661  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (716) 263-1304  
: TELEFAX: (716) 263-1600  
: INFORMATION FOR SEQ ID NO: 2:

: SEQUENCE CHARACTERISTICS:  
: LENGTH: 1838 amino acids  
: TYPE: amino acid  
: STRANDEDNESS:  
: TOPOLOGY: linear  
: MOLECULE TYPE: protein  
: US-09-596-784-2

Query Match 100.0%; Score 9448; DB 19; Length 1838;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1838; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MEKLSLGEHKAHVHTAAHNPVGHGVALQOGSSSSPQNAASAAAEKGNKCKMPRIHQ 60  
Db 1 MEKLSLGEHKAHVHTAAHNPVGHGVALQOGSSSSPQNAASAAAEKGNKCKMPRIHQ 60  
QY 61 STAADGISAAHQKSPSLRGCLGTTKFSRPAQOGPTTHSKGATLRDLARDDGETQH 120  
Db 61 STAADGISAAHQKSPSLRGCLGTTKFSRPAQOGPTTHSKGATLRDLARDDGETQH 120  
QY 121 EAAAPDAARLTRSGGVYRRNDDMAGRPVMYKGGSEDKVPPTQOKRHQJLNNFQGMQRTMLS 180  
Db 121 EAAAPDAARLTRSGGVYRRNDDMAGRPVMYKGGSEDKVPPTQOKRHQJLNNFQGMQRTMLS 180  
QY 181 KMAHPASANAGDRLQHSPPHPPHSGSHHEIKKEPVGSKATTAHADRVETIAQEDDDSEFQ 240  
Db 181 KMAHPASANAGDRLQHSPPHPPHSGSHHEIKKEPVGSKATTAHADRVETIAQEDDDSEFQ 240  
QY 241 LHOORLARERENPPQPKLVATPISARFQKLTAVAESVLEGGTDTTQSPKPSMLKGS 300  
Db 241 LHOORLARERENPPQPKLVATPISARFQKLTAVAESVLEGGTDTTQSPKPSMLKGS 300  
QY 301 GAGVTPAVTLDKGLQAPDNPALNTLLKOTLGKTOHYLAHHAASDGSQHLNLDKNG 360  
Db 301 GAGVTPAVTLDKGLQAPDNPALNTLLKOTLGKTOHYLAHHAASDGSQHLNLDKNG 360  
QY 361 HLFDIKSTATSYSLVHNSHPGEIKGLAQAGTGSVSDGKSGKISLGGTOSHNTKMLSQ 420  
Db 361 HLFDIKSTATSYSLVHNSHPGEIKGLAQAGTGSVSDGKSGKISLGGTOSHNTKMLSQ 420  
QY 421 PGEAHESLTGIWQHAGAPGOSIRLHDDKIHILHPELVGWQADKTHSOLROAD 480  
Db 421 PGEAHESLTGIWQHAGAPGOSIRLHDDKIHILHPELVGWQADKTHSOLROAD 480  
QY 481 GKLYALKDNRTLQNLSDNKSSEKLVKISYSDVQRCQVAILTDPGRHKMSIMPSLDAS 540  
Db 481 GKLYALKDNRTLQNLSDNKSSEKLVKISYSDVQRCQVAILTDPGRHKMSIMPSLDAS 540  
QY 541 PESHISLSLHFADAHOGLLHGKSELEAQSVASISGRILVADSEKGLFSAAIPKQGDGNE 600  
Db 541 PESHISLSLHFADAHOGLLHGKSELEAQSVASISGRILVADSEKGLFSAAIPKQGDGNE 600  
QY 601 KMKAMPQHALDEHFGHDOISGFFHDDHGLNALVKNFQOQACPLGNHDPHFGHNL 660  
Db 601 KMKAMPQHALDEHFGHDOISGFFHDDHGLNALVKNFQOQACPLGNHDPHFGHNL 660  
QY 661 DALVIDNQLGLHHTNPEPHEILDMGHLSLALQEGKLHYFDQLTGKWTGAESDCKQLKKG 720  
Db 661 DALVIDNQLGLHHTNPEPHEILDMGHLSLALQEGKLHYFDQLTGKWTGAESDCKQLKKG 720  
QY 721 LDGAAYLLKDGVEVKRLNTNQSTSSIKHGTENFSLPHVRNKPEPGDALQGLNKDDKAQ 780  
Db 721 LDGAAYLLKDGVEVKRLNTNQSTSSIKHGTENFSLPHVRNKPEPGDALQGLNKDDKAQ 780  
QY 781 AVIGVNYKYLALTEKGDIRSFQIKPGTQQLERPAQTLSREGISGELKDIHVHDKONLALT 840  
Db 781 AVIGVNYKYLALTEKGDIRSFQIKPGTQQLERPAQTLSREGISGELKDIHVHDKONLALT 840  
QY 841 HEGEVFHQPREAWQNGAESSSWHKALPOSSEKLSLDMSEHKPIATFEDGSHQHLKAG 900  
Db 841 HEGEVFHQPREAWQNGAESSSWHKALPOSSEKLSLDMSEHKPIATFEDGSHQHLKAG 900





QY 1506 PTFNLGVG---AGANLTAALGAHVSSTH---EGKPVGIFPAFTSNVSAALALDNRTSQ 1558  
 DB 1829 NEKQSGDGRHSGSRHEHASSRADSRHSGVGGQSSG---PRTSRNQSSVSQDS-DSQ 1884  
 QY 1559 SISLELKRAEPVTSNDISELTSTL-----GRHFKD-----SATTKMLAALKELD 1602  
 DB 1885 GHSEDSERWSGSASR--NHLGSAWEQSRDGRHSGHEDRAGHGHASDSSRQSGTRHTE 1942  
 QY 1603 DAKPAQLHLLOHFAKDVGVDEREAVRNKLVIRQAADS--HS-MELGSASHST 1659  
 DB 1943 SSSRGQ---AASSHEQARSAG-ERHSGHQL-----OSADSSRHSGHGHQAS----- 1987  
 QY 1660 YNNLSRINDGIVELHKLHFDALPASSAKRLGEMMNNDPALKDIIKOLQSTPFSASVS 1719  
 DB 1988 ----SAVRDSG-----HRGYSQASDSE--CHSESD-----TQSVS 2019  
 QY 1720 MELKDLGRLQTEKAILDKGVREVEGVLFODRNRLRVKSVSVSQSVSKSGCFNTPA 1775  
 DB 2020 AOKGAGPHOOSHRESARGQSGES-----SGRSGSLYQVSTHESQSESTHGQSAPS 2069

RESULT 9

US-09-450-969-4098  
 ; Sequence 4098, Application US/09450969  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lynn Doucette-Stamm et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
 ; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: PATH99-09A  
 ; CURRENT APPLICATION NUMBER: US/09/450,969  
 ; CURRENT FILING DATE: 1999-11-29  
 ; NUMBER OF SEQ ID NOS: 7544  
 ; SEQ ID NO 4098  
 ; LENGTH: 10203  
 ; TYPE: PR1  
 ; ORGANISM: S.epidermidis  
 US-09-450-969-4098

Query Match 2.7%; Score 257.5; DB 18; Length 10203;  
 Best Local Similarity 17.7%; Pred. No. 4.7e-08;  
 Matches 396; Conservative 319; Mismatches 805; Indels 723; Gaps 90;  
 QY 33 SSSSPQNAASL-AEGRNKRMPRIHOPSTAAAGISAAHQQKKSFSLRGCLGTHKKFSRS 91  
 DB 6262 ATQSVQNAEQALHGAELKNQDK-----QTSSTELDLTLTDAQRE-KLREQINTNSRSD 6316  
 QY 92 APQ--GQPGTTHSKGATLRLDRLARDG-----ETOHEAAAPDAAR----- 129  
 DB 6317 IKQIEQAKALNDAMKLUKEQVQAKDGVHANSDYTNEDSAOKDAYNNALKOAEIINNSS 6376  
 QY 130 -----LTRSGGKYKRNMDDMAGRPVKGSGEDKYPTQOKRHLNFGOMRTMLSK 181  
 DB 6377 NPNLNAQDITNALNNIKOQADNLHGAOKLQ---QDKNTTNOAIGNLNLHNPQOKDALIQ 6432  
 QY 182 MAHPASA-----NACDRLOHSPPHIPGSHHEIK----- 209  
 DB 6433 ATNGATSRDQVAEKLEAEALDEAMKLEDDQVNDQDQISNSPFTNEDSDKOKTYNDKIQ 6492  
 QY 210 --EEPVGSTSKAT-----TAHADREVEIAQEDDDSEFFQQLHOORLARE 249  
 DB 6493 AKELIINTSPTLDKQKIADTLQNKDAVNNLHGQD-KLAGSKODANNOLNHLDDLTEE 6551  
 QY 250 RENPPQPKLGVATPISARFQPKLFAVESVLEGTDTTQSPILKPSMLKSGSAGVTPLAY 309  
 DB 6552 QKN-----HFKP-----LINNADTRDEVNK----- 6571  
 QY 310 TLDKGLQLAPNPALNTLLKQTLGKQTHYLAHASSDGSQHLILLDNKGLFLDIKSTA 369  
 DB 6572 -----OLEIAKOLGMDSTLHVKINDKQIQHLSNYINADNDKKQNYD-----AIK 6618  
 QY 370 TSVSVLHNSHPGEIKGLAQAGTGSVVDGKSGKISLGSCTQSHNKTMLSOPGEAHRSL 429

DB 6619 EAEEDLHN-HPDPLDHKALQ-----DLNKKIDQAHNEL- 6650  
 QY 430 TGIWOHPAGAAR-----PQGESIR---LHDDKIHILHLPGLVWQS 466  
 DB 6651 -----NGESRFPQALDNALNIDISLNVQRTVKDNINHVTTLLESIAQEL----- 6698  
 QY 467 ADKDTHSQLSROADGKLYALKD-----NRTLONLSDNKSKSEKLVDKI----- 508  
 DB 6699 -----QKAKELNDAMKMRDSIMNOEQIRKNSNYTNEDLAQONAYNHAVDKINHIGE 6751  
 QY 509 -----KSYSDVORGQVAIILTDTPGRHKMSIMPSLD 538  
 DB 6752 DNATMDPQIITKQATODINTAINGLQDKLODAKTDAKQITNTGLTEPQKQAL----- 6806  
 QY 539 ASPESHISLSLHFADAHQGLLHGK---SELEAOSVALSHGHLVVAD-----SEGLKF 587  
 DB 6807 ---ENIINQOTSANAVAKQLSHAKFLNGKMEELKVAVAKASLVRSNYSYINEDYSEKAY 6863  
 QY 588 SAAIPKQGDGNE-LMKAMPQHALDEHFGHDHJSGFFHDDHGLNALVKNKFRQOHACP 646  
 DB 6864 EQAIAK---GOEIIINSENNTTISTDINRTIOEINDAEQNUHGD-----NKLQOAEI- 6913  
 QY 647 LGNDHOFHPGWNLTDALVIDNQLGLHHTNPEPHEILDMHGLSLALOFGLUHYFDQLTKG 706  
 DB 6914 AKNEIQLNDGLNSAQITKLIODIGRTTTPKPAVTQKLEAKAINQAMOOKOSIADK----- 6969  
 QY 707 WTGAESCKQLKGLDGAAYLLKDGVEKRLNIN-----QSTSSIKKHGTENVFSLPHYRN 760  
 DB 6970 -----DATLNSNYLNEDESEKKLAYDNAVSAQAEQLINQNDPTMDSINLOAITQ 7018  
 QY 761 KP-EPGDALQGLNKDDKAQAMAVGVNKYLALTEKGD-----IRSFQIKPFTQOOLRPA 813  
 DB 7019 KVIQAKSLHGANKLAQNOADSNLIINGSTNLNDKQKQALNDLNHQAOTKQOVAEIIAQA 7078  
 QY 814 QTLREGISGLKDIHVD-----HKONLYALTHEGEVFFHOPREAWONGAESSW----- 862  
 DB 7079 NKLNNE--MGTLLKTLVEQSVNVHQQSKY-INEDPQVQNIYNDISQIKGREILNGTTDDVLN 7135  
 QY 863 -HKLALPQSESKLAKSDMSHEKPIATFEDGSOH-----OLKAGGMHAYAAPERG 911  
 DB 7136 NKIADAATONILHTKNDLHGUQOKLOKAOQADATNELNYLNLNNSORQSEHDEINSAPSRT 7195  
 QY 912 PLAVGTSGSOTVFNLMQGVKVIPIPSGLTVKLSAQGTGGMGTGABGRKVSSEFSEIRAY 971  
 DB 7196 EVSNDLNHAKAL-NEAMQLENEVALENSVK-KLSDFIN-----EDEAAQNEYSNALQA 7248  
 QY 972 --AFNPTMST---PRPIKNAAYATOHQWQREGKLPFLYEMOGALIKOLD---AHN----- 1018  
 DB 7249 KDIIINGVPSSTLDKATIEDALLELQNAHRESLHGEOQLQEAQKQVAEIDNLQALNPGQVL 7308  
 QY 1019 -----VRHNAPOPDLQSLKETLDLGEHGAELLNDMKRFRDELEQSATRSVTVLGHQGVYL 1073  
 DB 7309 AEKTLVNOASTKPEVQEL-----QKAKELNEAMKALTEINKK-----EQIK 7351  
 QY 1074 KSNGEINSEFPKSPKALVQSPFN--VNRSGQ-----DLSKSQOAVHATPPSAESKL 1123  
 DB 7352 ADSRVVNAD-----SGLQANTYNSALNYGSIQIATTPPELAKNDVINRATQIKTAENL 7405  
 QY 1124 --QSMGLHFVSAG-VDMSHQKGEIPLGRQDPNDKLTALTKSLRLIDTVTIGELHELADKA 1180  
 DB 7406 NQOSKLAELAKSDGNSIEHLQ-----LHOSOK-----DKOHDNLNOA 7443  
 QY 1181 KLVSDBKPDADQI---KOLROQFDTLREKRYESNPVKHYTD-----MGFTPHKALEA 1229  
 DB 7444 QT-----KQOVDIVNNSKQDLSNMNLOQIIVNNDNTVKNSDFINEDSSOODAYNHAIOA 7499  
 QY 1230 NYDAVKAFTNAFKKEH-----HCVNLTR----- 1253  
 DB 7500 AKDLTAHTPTMDKNOIDAENIKOALNDLHGSNKLSEDKKEASPOLQNLNLSLTNGOKD 7559  
 QY 1254 TVLESQGSSEL-----AKKLNKTLSL-----DSCESMSFSRSYGGGVSTVFVPT 1298

```
Db 7560 TIINHIFSAPTSOVGEKIASAKOLNNTMKALRDSIADNNEILOSKYFNEDSEQ----- 7614
QY 1299 LSKKVPVPVPIPGAGITLDRAYNLSFSRTSGGLNVSFGRDGVSGNIMVATGHDVPMYTG 1358
Db 7615 -----QNAVNAVNAKAKNIIN-----DQPTVMANDEIQSVLNE 7648
QY 1359 KVTYSAGN-----ASDWLSAKHKISPLDRIGRAVSGTLOQTQNSLK----- 1399
Db 7649 VKOTKDNLHGDKQLANDKTDQAATLNALNLYNQAQRNLETQVQNSNSRPEVQKVVQLAN 7708
QY 1400 -----EKLTEDELPGFTHGLTHGLTPAEALLQKIEHQMGKSLTFSVDTS--ANLD-- 1450
Db 7709 QLDAMKKLDDALTG-----NDAIKOTSNY-INEDTSQOVNDEY 7747
QY 1451 LRAGINLNEGSKPNGVTARVSA-----GLSASANLAAGSRERSTTSQFGSTTS 1500
Db 7748 TDRGKNIVAEOTNPNNPNTINIADKITEAKNDLHGQVQKLKQAOQSINFINQMTGLNQ 7807
QY 1501 ASN-----NRPTFLNGV--GAGANLTAALGVVAHSSTHEGKPVGIF 1538
Db 7808 AQOEQLNQEIQOTQTRSEVHQVINKAQLNDSMNTLRQSTIDEHEVKQTSNYINETVGNQ 7867
QY 1539 PAFSTSNVSAALALDNRTSOSI--SLELKRAPVTSN-DISELTSTLCKHKFKDSATTKML 1595
Db 7868 TAYNNA-VDRVQIINQTSNPTNPNLEVERA-----TSNVKISKDALHGERELNDKNSKTF 7923
QY 1596 AA--LKELDKAPAEQLHILQOHF-----SAKDYYG----- 1624
Db 7924 AVNHLNLAQAQKEALTHEIEQATIVSOVNNIYKAKALNNDMKLKDIIAQQDNVQSN 7983
QY 1625 -----DREYAVRNKLLVITRQQAAD--SHSM 1649
Db 7984 NYINEDSTPQNNYNDTINHAOSIIDOVANPTMWSHDEIENAINNIKHAI-----NALDGEHL 8040
QY 1650 ELGSASHSTYNNLSRIN--NDGIVELLKHFDALPASSAKRLGEMMNDPALKDIK 1706
Db 8041 OQAKENANLLINSNDLNAPODAINRLVNE-----AQTRKVAEQLSQAALNDAMK 8093
QY 1707 QLOSTPFSASVSMELK-----DGLREOTEKAILDGKVGREEVGVLFQ-----DRNNLRV 1756
Db 8094 HLRNSIQNSQVROESKYINASDAKKEQYNHAV-----REVENIIEQHPTLDKELIK 8147
QY 1757 KSVSVSQSVKSEGTPTALLGTSNSAAMS-----MERNIGTINFKYGODONTPTRETFLE 1812
Db 8148 LTDGVNQA--NNDLNGVELLDADQONAHQSIPTLMHLNQAQNALNEKINNNAVTRTEVA 8204
QY 1813 GGIAQANQVASALTDLKKEGLE 1835
Db 8205 AIIGQAK-----LLDHAMENLE 8221

RESULT 10
US-08-827-356-5564
; Sequence 5564, Application US/08827356
; GENERAL INFORMATION:
; APPLICANT: George H. Shimer, Jr.
; APPLICANT: George H. Miller
; APPLICANT: Roberta S. Hare
; APPLICANT: Karen J. Shaw
; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS RELATED
; NUMBER OF INVENTION: COMPOSITIONS AND METHODS
; NUMBER OF SEQUENCES: 5574
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schering-Plough Corporation
; STREET: 2000 Galloping Hill Road
; CITY: Kenilworth
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07033-0530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
; SOFTWARE: PatentIn
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/827,356
; FILING DATE: 01-APR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/014,477
; FILING DATE: 01-APR-1996
; APPLICATION NUMBER: 60/016,743
; FILING DATE: 02-MAY-1996
; APPLICATION NUMBER: 60/020,016
; FILING DATE: 14-JUN-1996
; INFORMATION FOR SEQ ID NO: 5564:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3351 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: YES
; ORIGINAL SOURCE:
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1...3351
; US-08-827-356-5564
```

```
Query Match 2.6%; Score 246; DB 12; Length 3351;
Best Local Similarity 18.0%; Pred. No. 4.5e-08;
Matches 390; Conservative 299; Mismatches 797; Indels 676; Gaps 94;
```

```
QY 11 KAAVHTAAHNPVGHVALQOGSSSSSPQ--NAAASLAAGKRGKMPRIHOPSTADG- 66
Db 214 KQAYTDAYNAKNIV-----NGSPNVITNAADVTA--TORVNAETSNGD 259
QY 67 --ISAAHQKKSFSRLGCLGKTKFSRSAPQGPQGTTHSKGATLRDLARDGDTQ---HE 121
Db 260 TNLATAQQAAD-ALR-----QMTLSDAQKQSGITGIDSATQVTGVQ 301
QY 122 AAAPDAARLTRSGGVKRR--NMDDM-AGRPVVGKGGSEDKVPTQOKRHQLNFGCOMQT 177
Db 302 SVKDNATNLDNAMQLRNSIAKDEVKASQPYVDADT-----DKQNAVNTAVTSAEN 353
QY 178 MLSKMAHP-----ASANAGDRLOHSPPHIPGSHHKEEPPVSGTSKATTAHADR---VEI 229
Db 354 IINATSOPTLDPASVTAQAQNVNTNKTALNGAQN-----LANKQETTANINRLSHLNN 407
QY 230 AQEDD-----DSEFQOLHQ--QRL---AREENPPQPPKGLGVATPTSA 267
Db 408 AQKODLNTQVTNAPNISTVNOVKTKAEQLDQAMERLINGIQDKQVKQSVNFTDADPEKQ 467
QY 268 RFQPKLTAVAESVL---EGDTTTQSLPKPQSMKLGSGAGVTPLAVTLDKGLQLAPDNPP 324
Db 468 TAYNNAVTAENIINQANGINANGSQVE-----A 496
QY 325 ALATL--LKOTLCKDTHOYLAHASSDGSQHL--LLDNKGHLFDIKSTATSVYLHNSHPG 381
Db 497 ALSTVTTTQKALNGDRK---VTDAKNNAQNTLSTLDN-----LNNAQKG 537
QY 382 EIKGKLAQAGT-----GSVSVGDKSGKISLGSQT-----QSHN 414
Db 538 AVTGNINQAHTVAEVTAQIQTAEQLNTAMGNLKNLSNDKDTLGLSQNPADAPDEKKNAYN 597
QY 415 KTMLSQFGEAHRSLLTGIWQHPCAARPOG-----ESIRLH-----DDKTHI 456
Db 598 EAVRNAENILNKSTGTNPVKDQVEAAMNQVNTTKAALNGTQNLKAKQAHANTAIDGLSHL 657
QY 457 LHPELGWOSADKDTHSOLSRQADGKLYALKDNFTLNLSDNKSSEKLYDKIKSVYDQR 516
Db 658 -----TNAQKALKQLVQOSTTVAEAOGNEQKANNV--DAAMDKLRQSIADNATTKQ 707
QY 517 GOVAILTDTTPGRHKMSIMPSLDASPESHISLSLHFADAH-----OQLLHGKSELEAQS 569
Db 708 NQ-----NYTDASPNK-----KDAYNNNAVTAQGIIDQTTNPISLDP 743
```



; PRIOR APPLICATION NUMBER: US 09/036,081  
 ; PRIOR FILING DATE: 1998-03-06  
 ; PRIOR APPLICATION NUMBER: US 09/036,079  
 ; PRIOR FILING DATE: 1998-03-06  
 ; PRIOR APPLICATION NUMBER: US 09/035,913  
 ; PRIOR FILING DATE: 1998-03-06  
 ; PRIOR APPLICATION NUMBER: US 09/035,744  
 ; PRIOR FILING DATE: 1998-03-06  
 ; PRIOR APPLICATION NUMBER: US 08/827,356  
 ; PRIOR FILING DATE: 1997-04-01  
 ; PRIOR APPLICATION NUMBER: US 08/831,156  
 ; PRIOR FILING DATE: 1997-04-01  
 ; PRIOR APPLICATION NUMBER: US 60/014,477  
 ; PRIOR FILING DATE: 1996-04-01  
 ; PRIOR APPLICATION NUMBER: US 60/016,743  
 ; PRIOR FILING DATE: 1996-05-02  
 ; PRIOR APPLICATION NUMBER: US 60/020,016  
 ; PRIOR FILING DATE: 1996-06-14  
 ; NUMBER OF SEQ ID NOS: 7451  
 ; SEQ ID NO 4562  
 ; LENGTH: 3351  
 ; TYPE: PRF  
 ; ORGANISM: Staphylococcus aureus  
 ; US-09-611-529-4562

Query Match 2.6%; Score 246; DB 20; Length 3351;  
 Best Local Similarity 18.0%; Pred. No. 4.5e-08;  
 Matches 390; Conservative 299; Mismatches 797; Indels 676; Gaps 94;

```

QY 11 KAAVHTAAHPVGHVVALQOQSSSSSPQ---NAAASLAABGKRNKMPRIHQSTAADG- 66
Db 214 KOGATDAYNAKNIV-----NGSPNITNAADVTA--TORVNNAAETSLNGD 259

QY 67 ---ISAAHQOKKSFSLGCLGKTKFSRSPAPGQPGTTHSKGATLRDLILDGDTQ---HE 121
Db 260 TNLATAKQAKD-ALR-----QMTLSLDAQOKSITGQIDLSATQVTGVQ 301

QY 122 AAAPDAARLTRSGVKRR---NWDDM-ACRPVKGSGGKEDKVPQOKRRHOLNFGQMROT 177
Db 302 SVKDNATLNDNAMLNRSIAKNEKQASQYVDADT-----DKQAYNTAVTSAEN 353

QY 178 MLSKMAHP-----ASANAGDRLOHSPHPHPSHHETKEPVGSTSKATTAAHADR---VEI 229
Db 354 IINATSQPTLDPSAVTQAANOVNTNKTALNGAQN-----LANKQETTANIRLSHLNN 407

QY 230 AQEDD-----DSEFQQLHQ-QRL---ARENPQPPKGLGVATPISA 267
Db 408 AKQDLNTOVTNAPNISTVNVQVTKAEQLDQAMERLINGIDKQDVQKSVNETDADPEKO 467

QY 268 RFQPKLTAAVESVL---EGTDTTQSPKPSQMLKSGGAGVTPPLAVTLDDKGLQLAPDNP 324
Db 468 TAYNNAVTAENIINQANGTNANQSQVE-----A 496

QY 325 ALNLT--LKQTLGKDTQHYLAHSSDGSQHL-LDNDKGLHFDIKSTATSYSLVLSHSPG 381
Db 497 ALSTVTTTQKALNGDRK---VTDKANNANQTLSTLDN-----LNNAQG 537

QY 382 EIKGLAQAGT-----GSVSDGKSGKISLGSQT-----QSHN 414
Db 538 AVTGNINQAHVTAETQAIQTAQELNTAMGNLNSLNDKDTTLGSONFADADPEKKAYN 597

QY 415 KTMLSQFGEAHRSLLTGTIWOHPAGAAPQG-----ESIRLH-----DDKTHI 456
Db 598 EAVRNAENILNKSTGTNVPKDQVEAAMNQNVTAKAALNGTQNLKAKQHANTAIQGLSHL 657

QY 457 LHPELGVQWSADKTHSOLSRQADGKLYALDKNRTLQNLSDNKSSEKLVDKISYSDQR 516
Db 658 -----TNAQKALCOLVOQSTVAAEQNEQKANNV--DAAMDKLRQSTADNATTKQ 707

QY 517 GOVAILTDTGPRHRKMSIMPSLDASPESHISLSLHFADAH-----QGLLHCKSLEAQS 569
Db 708 NQ-----NYTDASPNK-----KDAYNNAVTTAQGIIDQITDTPNPSLDP 743
  
```

```

QY 570 VAISHGRLLVADSEGLFSAAIPKQGDGNELKMKAMPOHALDEHFGHDHQSIFFFHDDHG 629
Db 744 TVINQAAQGVSTSKNAL-----NGNE-NLEAAKQOAT-----OSLG 778

QY 630 QLNALVKNNFROQHACPLGNDHQFHGWNLTDLVIDNQLGLHHTNPPEHPEIILDMHIGS 689
Db 779 SLDTL--NNAQKQ-----AVTNOINGAHTVDEANOIKQNAQNLN 815

QY 690 LALQEGKLHYFDQLTGKWTGAESDCKQLKGLDGAAYLLKDGVEKRLININOST----- 742
Db 816 TAMGNLQKAIADKDATKATVNTFDQAQKQOAYNTAVTNAENIISKANGONATQTEVEQA 875

QY 743 -----SSIKHGTENVFSLPHVRNKPPEGDPALQGLNKDDKAQAMAVIGVNYKY 788
Db 876 IQQVNAAKOALNGNANVQHAKEATAL--INNSNDLQNAQKDKALKQOVQVQVATTVAGVN-- 931

QY 789 LALTEKGDIRSFOIKPGTQOLERPAQTLREGISGELKD-----IHVD-HKONLY----- 837
Db 932 -----NVKQTAQELNN-AMTQKQGIADKEQTKADGNFYNADSDKONAYNQAV 978

QY 838 -----ALTHEGEVFPHPREAWQGAESSWHKHLALPQSESKLSKSLDMSHEHKPIATFEDG 892
Db 979 AKAEALISGTPDVVVTPEI-----TAALNKVT--QAKNDLNG-----NTNLATAKQN 1024

QY 893 SOH-----OLKAGW-----HAYAAPERGPLAVGTSGSQTVPNRLMQGVKGV-IP 937
Db 1025 VQHALDQLPNLNOAORDEYSKOITQATLPVNVNATQQAATLNDAMTOLKQGTIANKAQIK 1084

QY 938 GS-----GLTVKLSAQGTGAGTKYSSKFSERIRAYAFNPTMTSTPRIPKNAAYATQH 992
Db 1085 GSENVHADTDKQATYDNAVTKA-----EELLKOTNPTMD-PNTIQOALTQKVD 1133

QY 993 GWQREGKLPIYE-----MQGALIKOLD-----AHNVHRHNAPOPD 1027
Db 1134 TNQALNGNQLADAKQADAKTTLGTLDLHNDQAQKALTQVEQAPDITATVNNVKNQNAQNLN 1193

QY 1028 -----LQSKLETL-----DLGEGHGAELL--NDMKRFRDELEQS 1058
Db 1194 NAMTNLNAQDKTETLNSINFDTADQAKDDYTNVSHAEIGILSKANGSNASQTEVEQA 1253

QY 1059 ATRSVTVLQGHQVGLKNGEIN-----SEFKPSGKALVQSF-----NVN 1098
Db 1254 MOR-----VNEAKQALNGNDNVQRAKRAKQVITNANDLNQAKDKALQOVQDAQVAVNV 1309

QY 1099 ---RSGQDLKSLQQAQVHATPPSAESKLSQMLGHFVSAGVD-----MSHQKGEIPLG 1147
Db 1310 TIKQTAODLNAQMTQLKQGIADKQDTKAN---GNFVNADTDKQNNAYNAVAHAEOII--- 1363

QY 1148 RQDPNDKTALTSLRLILDV--TIGEL---HEL---ADKAKLYSDHKPDADQ----- 1192
Db 1364 -SGTPNANVDPQOVAQALQOVNAQKGLNGNHNQVAKDNANTAIDQLPNLNQPKTALK 1422

QY 1193 -----IKQ-----LRQGFDTLREKRYESPVKHYTDMGFTPHNKALEANYDAY 1234
Db 1423 DQVSHAELVTGVNAIKONADALNNAMGTLLKOOIOANSQVPSVD--FT-----QADQKQ 1475

QY 1235 KAFINAFKKEHHGVNLTFRITVLES-----QGSAAELAKKLNTLLSLDS 1277
Db 1476 QAYNNAANOQAQIANGTPTPLVAPDVTYKAVTTNQAQKDALNGDEKLAQAKODALANLDT 1535

QY 1278 GESMSFSRSYGGGVSVFVFTLSKKVPVPIPGAGITLDRAYNLS-----FSRTSGGLNVS 1333
Db 1536 LRDLN-----QPORDALRNOINQAQALATVEQTKQNAQNVNTA 1573

QY 1334 FGR-DGVSIGNIMVATGHDVMPYMTGKTSAGNASDWL-----SAKHKISPDLRIGAV 1386
Db 1574 MGNLKGIAKQVTKASENVHDADVQKQATYNAVSAQEGLIINGTTPNTPDITRALT 1633

QY 1387 SGT-LOGTIONSLSKFLTEDELPGFTHGLTHGLTTPAELLQKIE-----HOMKQGS 1437
Db 1634 QVTDKNSLNGEAKLATEKQNAKDAVSGMTHLNDQAQKALQKQIDQSPETATVNOVKQ-- 1691
  
```





QY 1544 TNVSAALDNRTSISL-----ELKRAEPTVTSNDISELTSTGLKHFEDSATTMKLAAL 1598  
 Db 1263 -NLKHALAEKONTKRSVNYTDADQPKQAYDFAVTAQAEAITWANGSNAME---TQVAAAL 1318  
 QY 1599 KELDDAKPAEQLH-----LQOHFSKDVVGDREYEAVERNKLKLVITROQ-----A 1643  
 Db 1319 NOLNAQK--NDLNGDNKVAQAKESAKRALAS--YSNLLNNAQSTAATSQIDNATTAVGVT 1374  
 QY 1644 ADHSMELGSA-----SHSTTYNNLSRIINDGIVELLH-- 1676  
 Db 1375 AONTANELNTAMGOLONGINDONTVKQVNFDDAGKDDAYTN-AVTNAQGILDKAHQ 1433  
 QY 1677 ----KHFDAL-----PASSAK-RLGEMMN-NDPALKDIIKLOQSTP 1712  
 Db 1434 NMTRQAQVEAALNQVTTAKNALNGDANVRQAKSDAKANLGLTLHLNNAQKDLTSIEGAT 1493  
 QY 1713 FSSASVSMELK-----DGLREOTEKAILDGKVGREEVGLVFDQRNLRVKSVSQS--V 1765  
 Db 1494 TVNGVGVKTKAQLDQAGAMQRLQSAIAKNDQTKASENYIDADPTKKTAFDPAITQAESYL 1553  
 QY 1766 SKSEGFNTPALLGTSNSAAMSERNI-GTINFKYGDOQ-----NTPRRTFLBG 1813  
 Db 1554 NKDHGANKDKQAVEAQISQVSTENALNGDANLQRAKTEAIOAIDNLTHTLNPQKTALQ 1613  
 QY 1814 GIAQANPOVASALTDLKKEGLEMS 1838  
 Db 1614 QVNAA--QRVSGVTDLKNATSALNN 1636

## RESULT 13

US-60-253-625-2389  
 ; Sequence 2389, Application US/60253625  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Haselbeck, RH  
 ; APPLICANT: Ohlsen, KL  
 ; APPLICANT: Zyskind, JW  
 ; APPLICANT: Trawick, JD  
 ; TITLE OF INVENTION: Identification of essential genes in  
 ; TITLE OF INVENTION: Staphylococcus aureus, Pseudomonas aeruginosa, Klebsiella  
 ; FILE OF INVENTION: pneumoniae and Salmonella typhimurium  
 ; FILE REFERENCE: ELITRA.017PR3  
 ; CURRENT APPLICATION NUMBER: US/60/253.625  
 ; CURRENT FILING DATE: 2000-11-27  
 ; NUMBER OF SEQ ID NOS: 2768  
 ; SOFTWARE: FASTSEQ for Windows Version 4.0  
 ; SEQ ID NO 2389  
 ; LENGTH: 2434  
 ; TYPE: PRF  
 ; ORGANISM: Staphylococcus aureus  
 ; US-60-253-625-2389

Query Match 2.6%; Score 242; DB 23; Length 2434;  
 Best Local Similarity 18.2%; Pred. No. 5e-08;  
 Matches 311; Conservative 238; Mismatches 656; Indels 500; Gaps 73;  
 QY 493 QNLSDNKSEKLVDKI-KSYSDVGQGVAILTDPGRHKMSIMPSLDASPEHSISLSLHF 551  
 Db 73 QNLAQAKTN--ATNTINNAHDLNQKQDALKTVNNARQVS-----DANNVQHTATELNS 125  
 QY 552 A-DAHQGLLHKSELEAQSVAISHGLRVADSE----- 583  
 Db 126 AMTALKAATADKERTKAS-----GNTVNAQDOEKRAQYDSKVTNAENIISGTPNATLTVN 179  
 QY 584 -----GKLSAIPKQDGNELMKAMPQALHDEHFGHDOISGFEHDDHQLNAL--- 634  
 Db 180 DVNSAASQVNAKALNGDNN---LRVAKHA-----NNTIDGLAQLNNAQAKLKEQ 229  
 QY 635 -----VKNFPROQHACPLG-----NDHPFHQGNLTD----- 661  
 Db 230 VQSATTIDGVQTVKNSSQTLNTAMKGLRDSIANEATIKAGONYTDAFPNNRNEYDSAVTA 289

QY 662 ALVIDNQLGLHHTNP--EPHEILDMLGSLALQEGKLUHYFDQLTKGWTGAESDCQK---- 716  
 Db 290 AKAIINOT-----SNPTMEPNTITQV--TSQVTTKEQALNGARNLAAQAKTTAKNNLNLT 343  
 QY 717 -----LKKGLDGAAYL--LKDGEVKRLINOSTSSIKHGTENVFSLPHVVRNKEPEGD 766  
 Db 344 INNAQRDALTRSIDGATTVAGVNOETAKATELNNAMHSLQNG-----INDTOTQK 394  
 QY 767 ALQGLNKDDKAQAMAVIGVNKYLALTEKQDIFRSQIKPGTOOLERPAOTLSREGISGELK 826  
 Db 395 TOKYLDAPESKKSAYDAQVANAAKAILTKASQONVDKAAVEQALQNVNSTKTALNGADKLN 454  
 QY 827 DIHVDHKONLYALTHEGEVPHOPREAWQN-----GAESSWHKL-----ALPQSES 873  
 Db 455 EAKAAAKQTTGTLTH-----INNAQRTALDNEITQATNVGVNTVRKAKAQLDQMGMOLETS 511  
 QY 874 LKSLDSHSEHKPIATFEDGSOHLK-----AGCWHAYAAPERGPPLAY----- 915  
 Db 512 IRDKDITTLQSONYODADDAKRTAYSOAVNAATILNKTAGGNTPKADVEMAQVTOANT 571  
 QY 916 GTSQSQTIVNRLMQGVKGVIPGSGGLTVK---LSAQ---TGGMTGAEG----- 957  
 Db 572 ALNGIONL--DRAQQAANTAITNASDLNTKQKALKAQVTSAGRVSAANGVEHTAELNTA 630  
 QY 958 -----RKVSSK-----FSERIRAYAFNPTMTSTPRP-----IKNAA 987  
 Db 631 MTALKRAIADKAETKASGNVYNADANKRQAYDEKVT-AENIVSGTPTPTTTPADVNTAA 689  
 QY 988 -----YATQHGQWQREGKPLYEMOGALIKOLDADHVRHNAPOPDLOS 1030  
 Db 690 TOVTNAKTQNLNGHNLEVAKONANTAIDGLTSLNGPQKAKLKE---QVGQATTLPNVQT 745  
 QY 1031 KLETLDLGEHGAELLND-MKFRFDELQOSATRSVTYVIGOHQGVLSKNGEINSEKPSGPK 1089  
 Db 746 VRD-----NACTLNTAMKGLRDSIANEATIKA---GONYTDSQNKQOTDYSNAVTA 795  
 QY 1090 ALVQSFNVNRSQDLSKSLQ-QAVHATPPPSAESKLSQSLMGHVFVSAGVDMSHQKEIPLGR 1148  
 Db 796 AIT-----GQTTSPSMNAQEIQAQVTAQKQALNG----- 827  
 QY 1149 QRPDNDKTALTKSRLLDVTIGELHELADKAKLVSDHKPDA-----DOIQLR 1197  
 Db 828 --QENLRTAQTNK-----OHLNGLSD-----LTDQAQDAVKRQIEGATHVREVTQAQ 873  
 QY 1198 QOQFD-----TLREKRYESNPVKHYTDMGFTHNKALAEANYDAVKATINAKKHHGVNL 1250  
 Db 874 NNADALNTAMTNLKNIGIQDONTIK---QGVNFTDADEAKRN---AYTNAVTAQEQILNK 926  
 QY 1251 T-----TRTVLESQSAE-----LAKKLNLTLLSDGESMSFSRSY 1287  
 Db 927 AQGPNTSKDGVETALENVORAKNELNGNVANAKTTAKNALNLTINNQAKEALKSQI 986  
 QY 1288 GGVSVTFVFTLSKKVPVPIPGAGITLDRAYNLSFRTSGGLN-----VSEGR 1336  
 Db 987 EGATTVAGVNOVS-----TTASELNTAMSNLNGINGINDEAATKAAQKYTDADR 1033  
 QY 1337 DGVSGNINMVAICHVDVMPYMTGKTSAGNASDWL-----SAKHKISPDLRICAAGVTLQ- 1391  
 Db 1034 EKQATYNDAVTAAKTLLDKTAGSNDNKAQVEQALQVNTAKTALNGDERLNEAKNTAKQ 1093  
 QY 1392 -----GTLQNSLKFTEDELPGFIHGLTHGLTTPAELLQKGIHQMKQKGLTFSDPTS- 1446  
 Db 1094 VATMSHLTDAQKANTLSQIESGTTVAGVOIQANAGTLQDAM-NQLRQSIASKDATKSSE 1152  
 QY 1447 ----ANLDRLAGIN-----LNEDGSKPNGVTARVS-----AGLSASANLAAG 1484  
 Db 1153 DYODANADLQONAYNDAVTNAEGIIISATNPMPNPDITNOKASQVNSAKSALNGDEKLA 1212  
 QY 1485 SRSRTTSGQFSTTSASNNRPTFLNG-VGAGANLTAALGVASHSSHEGPKVGFPAFTS 1543  
 Db 1213 ---KOTAKSDICRLTDLNNAQRTAANAQVDPAPNLAATAAKNKAATSLNTAMG----- 1262  
 QY 1544 TNVSAALDNRTSISL-----ELKRAEPTVTSNDISELTSTGLKHFEDSATTMKLAAL 1598

Db 1263 -NLKHALAEKDNTRSVNYTDADQPKQAYDTAVTQAEAITNANGSNANE---TOVOAAL 1318  
 QY 1599 KELDDAKPAEQLH-----ILQOHFSAKDVGVGBERYEAVNLKLVIRQO-----A 1643  
 Db 1319 NOLNOAK--NDLNGDNKVAQAKESAKRALAS--YSLNNAAGSTAATSDIDNATVAGVTA 1374  
 QY 1644 ADHSHMELGSA-----SHSTTYNNLSRINNDGIVELLH-- 1676  
 Db 1375 QONTANELNTANGQLONGINDQNTYKQVNFDDAQGKKDAYTN-AVTNAOGLDKAHQO 1433  
 QY 1677 ----KHFDAAL-----PASSAK-RUGEMMN-NDPALKDIKIQLOSTP 1712  
 Db 1434 NMTKAQVEAALNQVTTAKNALNGDANVRQAKSDAKANLGLTTHLNAQKQDLSQIEGAT 1493  
 QY 1713 PSSASVSMELK---DGLREOTEKAILDGKVGREEVGLFODRNLRKVSQSVSQS---V 1765  
 Db 1494 TVNGVGVKTKAODLDGAMQRLQOSALANKDQTKASENYIDADPTKTAFDNATQAESYL 1553  
 QY 1766 SKSEGFNTPALLIGTSNSAAMSERNI-GTINFKYGQDO-----NTPRFTLEG 1813  
 Db 1554 NKDHGANKDKQAVEAQIQSVTSTENALNGDANLQRAKTEAIOAIDNLTHTLNPQKLTAKQ 1613  
 QY 1814 GTAQANPQVASALTDLKKEGLEKMS 1838  
 Db 1614 QVNAA--QRVSGVTDLKNATSLSLN 1636

RESULT 14

US-60-257-931-3392  
 ; Sequence 3392, Application US/60257931  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Haselbeck, R  
 ; APPLICANT: Ohlsen, KL  
 ; APPLICANT: Zyskind, JW  
 ; APPLICANT: Trawick, JD  
 ; APPLICANT: Wall, D  
 ; TITLE OF INVENTION: Genes identified as essential in Staphylococcus aureus, Salmonella  
 ; FILE REFERENCE: typhimurium, Klebsiella pneumoniae and Pseudomonas aeruginosa  
 ; CURRENT APPLICATION NUMBER: US/60/257,931  
 ; CURRENT FILING DATE: 2000-12-22  
 ; NUMBER OF SEQ ID NOS: 3592  
 ; SOFTWARE: FASTSEQ for Windows Version 4.0  
 ; SEQ ID NO 3392  
 ; LENGTH: 2434  
 ; TYPE: PRF  
 ; ORGANISM: Staphylococcus aureus  
 ; US-60-257-931-3392

Query Match 2.6%; Score 242; DB 23; Length 2434;  
 Best Local Similarity 18.2%; Pred. No. 5e-08;  
 Matches 311; Conservative 238; Mismatches 656; Indels 500; Gaps 73;

QY 493 ONLSDNKSSEKLVDKI-KSYSDVORGQVAILTDPFRHKMSIMPDLSPESHISLSLHF 551  
 Db 73 QNLAQAKTN--ATNINAHDLNOKQDKALTKQVNNARVS-----DANNVQHTATELNS 125  
 QY 552 A-DAHOGLLHGSKELEAQSVAISHGRVLVADSE----- 583  
 Db 126 AMTALKAADIADKERTKAS-----GNVYNADOEKROAYDSKVTNAENIISGTPNATLTVN 179  
 QY 584 -----KGLPSAAIPKQDGNELKWKAMPOHALDEHGHDOHISGFFHDDHGNAL--- 634  
 Db 180 DVNSAASQVNAAKTALNGDNN---LRVAKHA-----NNTIDGLAQLNNAQAKLKEQ 229  
 QY 635 -----VKNNPQROHACPLG-----NDHQHPGWNLT----- 661  
 Db 230 VOSATTLDCGVQTVKNSGTLTAMKGLRDSIANEATIKAGQNYTDASPNRNEYDSAVTA 289  
 QY 662 ALVIDNOLGLHHTNP--EPHEILDMGLHGLSALQEGKLHYFDQLTKGWTGAESCKQ--- 716

Db 290 AKAILINOT-----SNPTMEPTITQV--TSQVTTKEQALNGARNLAQAKTTAKNNLNLLTS 343  
 QY 717 -----LKGLDGAAYL--LKDGEVRLNINQSTSSIKHGTENVFSLPHVRNKEPEPD 766  
 Db 344 INNAOKDALTRSDGATTVAGVNOETAKATELNNAMHSLONG-----INDETQTKQ 394  
 QY 767 ALOGLNKDDKAAAMAVIGVKNYLALTEKDIRSFQIKPGTQOOLERPAQTLRSREGISGEIK 826  
 Db 395 TOKYLDAPESKSAVDQAVNAAKAILTKASQGVNDKAAVEQALONVNSTKTALNGDAKLN 454  
 QY 827 DIHVDBKQNLIALTHEGEVPHQPREAWQN-----GAESSWHKL-----ALPOSEK 873  
 Db 455 EAKAAAKQTLGLTH---INNAQRTALDNEITQATNVEGVNTVKAKAQQLDCAMGOLETS 511  
 QY 874 LKSLDMSHHKKPIATFEDCSQHLK-----AGWHAYAAPERGLAV----- 915  
 Db 512 IRDKDTTLOSQNYQDADDAKRTAYSOAVNAAATILNKTAGGNTPKADVVERMAQVQANT 571  
 QY 916 GTSGSOTVFNRLMOGVKGVIPGSLTVK---LSAQ---TGGMTGAEG----- 957  
 Db 572 ALNGIONL--DRAKQAAANTAITNASDLNTKQEKAKQVTSAGRVSAANGVEHTATELNTA 630  
 QY 958 -----RKVSCK-----FSERIRAYAPNPTMSTPRP-----IKNAA 987  
 Db 631 MTALKRAIADKAETKASGVNVNADANKROAYDEKVT-AENIVSGTPTTLPADVTNAA 689  
 QY 988 -----YATQHGWOEGREGLKPLYEMOGALIKOLDAHVNRHNAPOPDLOS 1030  
 Db 690 TQVTNAKTQNGNHNLEVAKONANTAIIDGLTSLNGPOKAKLKE-----OVGQATTLPNVOT 745  
 QY 1031 KLETDLGEGHCAELLND-MKRFRDELEQSATRSVTVLGOHGVKLSNGEINSEFKPSGK 1089  
 Db 746 VRD-----NAOTLNTAMKGLRDSIANEATIK---QGVNTDASQNKQTDYNSAVTAAK 795  
 QY 1090 ALVOSFNVNRSGQDLKSKLO-QAVHATPPSAESKLOSLMGLHFYSAGVDMSHQKGEIPLGR 1148  
 Db 796 AII-----GOTTSPSMNAQETNOAQDQVTAQKQALNG----- 827  
 QY 1149 QRDPNKDTALTKSRLILDTVTIGLHELADKAKLVSDHKPDA-----DQIKOLR 1197  
 Db 828 --OENLRTAOTNAK-----QHNLGLSD---LTDQAKDAVKRQIEGATHVNEVTQAO 873  
 QY 1198 QQFD-----TLREKRYESNVKHYTDMGPTHNKALLEANYDAVKAFINAFKKEHGVNL 1250  
 Db 874 NNADALNTAMTNLKNIGIQDQNTIK---QGVNFTDADEAKEN---AYTNVATQAEQILNK 926  
 QY 1251 T-----TRTVLESQSAE-----LAKKLNTLLSLDSESFSRSY 1287  
 Db 927 AOGPNTSKDGVETALENVQRAKNEELNGONVANAKTTAKNALNLTSLNNAQKALKSOI 986  
 QY 1288 GGGVSTVFVPTLSKKVPVPVPIPGAGITLDRAYNLSFSRTSGLN-----VSFGR 1336  
 Db 987 EGATTVAGVNVQS-----TTASELNTAMSLNQLNGINDEAATKAAQKYTDADR 1033  
 QY 1337 DGGVSGNIMVATGHDVMPYMTGKTKSAGNASDWL---SAKHKISPLDRIGAAVSGTLQ- 1391  
 Db 1034 EKQATYNDVATAAKTLTLDKTAGSNDNKAABEQALQRVNTAKTALMGDELNEAKNTAKQO 1093  
 QY 1392 ---GTLQNSLKFKLTEDELPCFIHGLTHGTLTTPAELLQKGIHQMGKSGKLTFSVDTS- 1446  
 Db 1094 VATMSHLTDAQKANLTSQIESGTTVAGVOGQANAGTLDOAM-NOLROSIAKSKDTSSE 1152  
 QY 1447 ---ANDLRAGIN-----LNEDGSKPNGVTARVS-----AGLSASANLAAG 1484  
 Db 1153 DYODANADLQAYNDVAVTNAEGIIISATNPNEMNPDTINOKASQVNSAKSALMGDEKLAAA 1212  
 QY 1485 SREKSTTSGQGSSTTSASNNRPTFLNG-VGAGANLTAALGVASHSTHECKPVGIFPAFTS 1543  
 Db 1213 ---KOTAKSDIGRLTDLNNAORTAANAEDQAPNLAATAAKNATSLNTAG----- 1262  
 QY 1544 TNVSAALADNRTSOSISL-----ELKRAEPTVSDISLTSTLKGHPKDSATTMLAAL 1598  
 Db 1263 -NLKHALAEKDNTRSVNYTDADQPKQAYDTAVTQAEAITNANGSNANE---TOVOAAL 1318

QY	1599	KELDDAKPAEOLH----	ILQOHFSKADVVDGERYEAVRNLKLVIRQQ-----	A	1643
DB	1319	NQLNOAK--	NDLNGDNKVAQAKESAKRALAS--	YSNLNNAAQSTAATSQIDNATTVAGVTA	1374
QY	1644	ADSHSMELGSA--	-----	SHSTTYNNLSRINNDGIVELLH--	1676
DB	1375	AQNTANELNTAMQOLQNGINDQNTVKQOVNFTDADQGGKKDAYTN-AVTNAAQILDKAHQO	1433		
QY	1677	----KHFDAAAL-----	PASSAK-RLGEMMN-NDPALKILIKOLQSTP	1712	
DB	1434	NMTKAQVVEAALNQVTTAKNALNGDANVROAKSDAKANLGTJLHLNNAQKQDLTSQIEGAT	1493		
QY	1713	FSSASVSMELK----	DGLREOTEKAILDGKVGREEVGLFODRNNLRKVSVSQS--	V	1765
DB	1494	TVNGVNCVKTKAQDLGDGMQRLOSALANKQTKASENIDADPTKKTAFDNRATQAESVL	1553		
QY	1766	SKSEGFNTPALLLGTSNSAAMSERNI-GTTFNKYQDQO-----	NTPRFTLEG	1813	
DB	1554	NKDHGANKQQAQVEAQTSVSTENALNGDANLQRAKTEATQOIDNLTLNTPORTALKQ	1613		
QY	1814	GIAQANPQVASALTDLKKLEGLEMS	1838		
DB	1614	QVNA--	QRVSGVTDLKNATSUNN	1636	

```

RESULT 15
US-60-269-308-4413
; Sequence 4413, Application US/60269308
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, RW
; APPLICANT: Ohlsen, KL
; APPLICANT: Zyskind, JW
; APPLICANT: Trawick, JD
; APPLICANT: Wall, D
; TITLE OF INVENTION: Identification of Essential Genes in Staphylococcus aureus, Pseudomonas aeruginosa, Klebsiella pneumoniae, Salmonella typhimurium, and Escherichia coli
; TITLE OF INVENTION: faecalis
; FILE REFERENCE: ELITRA.017PRS
; CURRENT APPLICATION NUMBER: US/60/269,308
; CURRENT FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 4774
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4413
; LENGTH: 2434
; TYPE: PRF
; ORGANISM: Staphylococcus aureus
US-60-269-308-4413

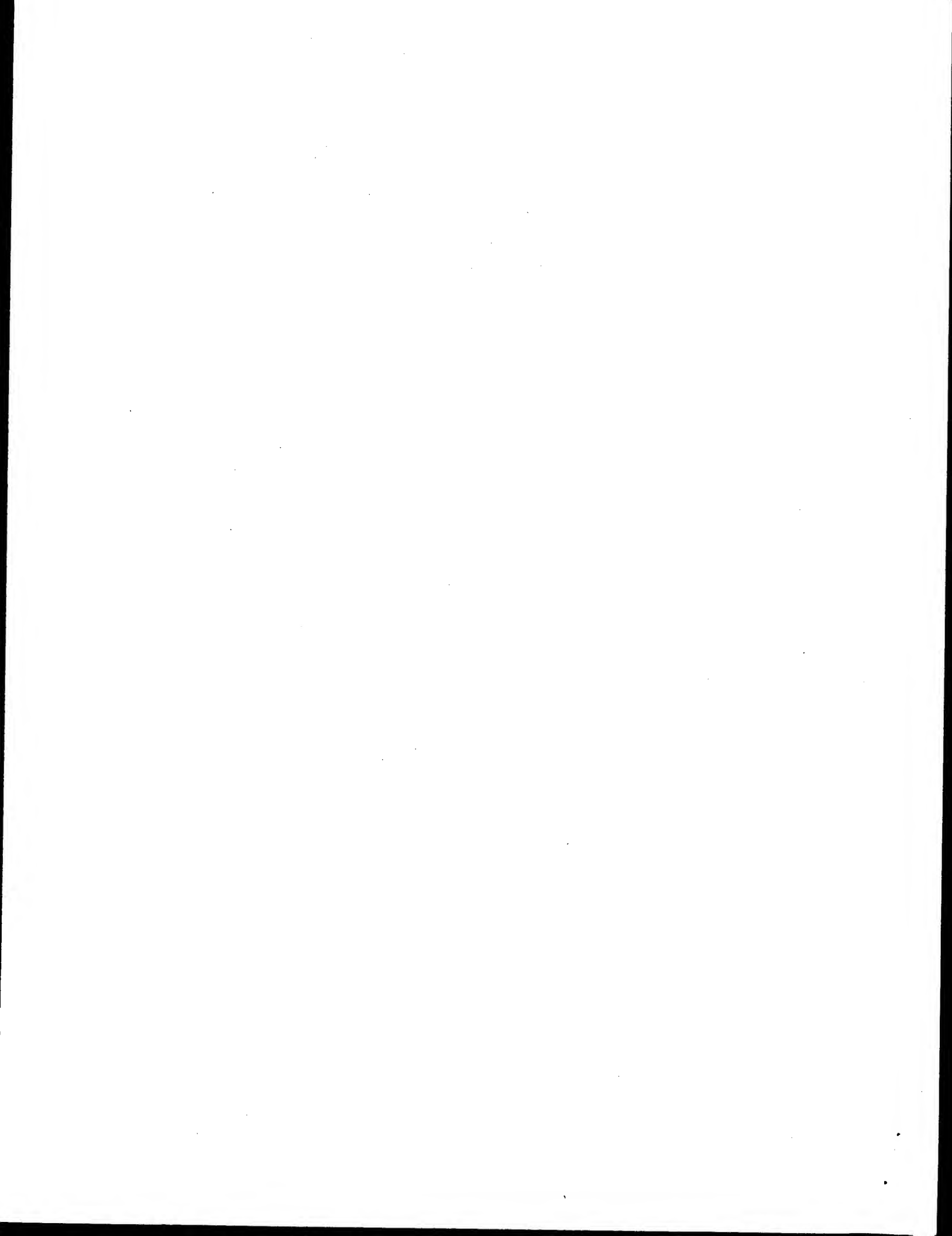
```

Query Match 2.6%; Score 242; DB 23; Length 2434;  
Best Local Similarity 18.2%; pred. No. 5e-08;  
Matches 311; Conservative 238; Mismatches 656; Indels 500; Gaps 73;

Qy	493	QNLSDNKSSEKIVDKI - KSYSYDQRGQVAILTDTPCRHKMSIMPDISDASPESHISLSLHF	551
Db	73	QNLQAQKTN--ATNTINNAHDLNKKDALKTOVNQAQRVSD--DANNYQHTATELNS	125
Qy	552	A-DAHOGLLHGKSELEAQSVASHGRLVYADSE-----	583
Db	126	AMTALKAAIADKERTKAS-----GNYNAQOEKQAYDSKVTVNAENIISGTPNATLVN	179
Qy	584	-----GKLFSAAIPKQGDGNEKMKAMPOHALDEHFDHDIQISGFHDDHQLNAL---	634
Db	180	DVNSAASQVNAAKTALNGDNN--LRVAKAHEA-----NNTIDGLAQLNNAQAKLKEQ	229
Qy	635	-----VKNNFRQQAACPLG-----NDHOFHPGNNLTID-----	661
Db	230	VQSATTLDGQVTVKNSSQTLNTAMKGLRDSITANEATIKAGONYTASPNNNNEYDSAVTA	289
Qy	662	ALVIDNQLGLHHTNP--EPHEILDMGHGLSIALQEGKLIHYFDOLTKWGTGAESCKO---	716
Db	290	AKAIINOT-----SNTMPEPNTIQV--TSQVYTTKEQALNGARNLQAQKTTAKNNLNILTS	343

[illegible]

Search completed: June 5, 2001, 18:20:30  
Job time: 356 sec







Db	946	KLFTGKDLKANNGATVSEDDGKD-----TGTLGVTAKTV---IDAVNK---SGWRVTG	994
QY	1246	HGVNLTTRTVLSEQSAEIAKLKNTLLSLDSEMSFSRSYGGVSTFVFPVTLSKKVPV	1305
Db	995	EGATAETGATAVNAGNAE-----TWTSGTSYNFK---NGNATTATV-----	1032
QY	1306	PVIPGAGITLDRAYNLSFRTSGGLNVSFCRDGVSNGTINMVGTHDVP-----YMTGKK	1360
Db	1033	-----SKONGNINVKY--DVNWGDGLKIGDDKKIVADTTTLTVTGGK	1072
QY	1361	TSAGNASOWLSAKHKISPDLRTICAAVSGTLOGTQLNSLKFKLTEDELPFIHGLTHGTLT	1420
Db	1073	VSPVAGANSVNNKKL-----VNAEGLATALNLSWTAKADK---YADGESEGE---	1118
QY	1421	PAELLQKGEHOMKOGSKLTFSDVTSANLLDRAGINLNEDGSK-----PNCVTVARVSA	1473
Db	1119	-----TDQEVKAGDKVTF-----RAGKNLKVQSEKDFYSLQDFTLTGLTISI	1160
QY	1474	GLSASANLAAGRSRSTTSGQFGSTTSASNRPFLFNGVGAGANITAAALGVASHSTHEGK	1533
Db	1161	TLGETAN---GRNDTGTWINKDGLTITLAN-----GAAAGTDSANGNTISVTKDGI	1208
QY	1534	PVGIPPAFTSNVSAAL-----ALONRTSOSISLELKRAPV-----TSNDISELTS	1580
Db	1209	SAG---NKEINVKSAITYKDTQNTADETQDKEFHAAVKVNAKEVFGKNGATVSAKTD	1265
QY	1581	TLGKHFK--DSATTKMLAALKELDDAK-----PAEQLHLIQOHFSAKD	1621
Db	1266	NGKHVTVIDVAEKVGVGDLEKDTDGKIKLVNDTDGNLLTVDATKGASVAKGFENA--	1323
QY	1622	VVGDERVEAVRNLK---KLVTROQAADSHSMELGSAHSSTTYNNLSRINNDGIVELLHKH	1678
Db	1324	VTTDATTAAQTANANERGVVVK-----GSGATATETDKKKVATVG-----	1364
QY	1679	FDAALPASSAKRLGEMMNDPALKDIIKOLSTPFPSSASVSMELKDGLREQTEKAILDGK	1738
Db	1365	-DVAKAINDAATFKVVENDDSATID-----DSPTDGANDALKAXDTL--TLKAGKNLK	1415
QY	1739	VGREEVGVLFODRNLRVKVSVSOSKSEGFNTPALLLGTSNSAAMSERNIGTINF-	1797
Db	1416	VKROGKNITFALANDLSVKATVSDKLS-----LGT-NGKNVNIITSDTKGLNFA	1463
QY	1798	---KYGOQNTPRFTLEGGIAQANQPVASALTD	1828
Db	1464	KDSKTGDDAN---IHLNG-----IASTLTD	1485

RESULT 2  
US-08-685-467-4  
; Sequence 4, Application US/08685467  
; Patent No. 6060059  
; GENERAL INFORMATION:  
; APPLICANT: St. Geme III, Joseph W.  
; APPLICANT: Barenkamp, Stephen J.  
; TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flehr, Hombach, Test, Albritton & Herbert  
; STREET: Four Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/685,467  
; FILING DATE: 22-JUL-1996  
; CLASSIFICATION: 424

```

/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/409,995
/ FILING DATE: 24-MAR-1995
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Silva, Robin M.
/ REGISTRATION NUMBER: 38,304
/ REFERENCE/DOCKET NUMBER: A-61053-2/RFT/RMS
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (415) 781-1989
/ TELEFAX: (415) 398-3249
/ TELEX: 910 277299
/ INFORMATION FOR SEQ ID NO: 4:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1912 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: unknown
/ TOPOLOGY: unknown
/ MOLECULE TYPE: protein
/ US-08-685-467-4

Query Match 2.2%; Score 209.5; DB 3; Length 1912;
Best Local Similarity 18.9%; Pred. No. 8.1e-06;
Matches 346; Conservative 226; Mismatches 653; Indels 609; Gaps 85;

QY 236 SEFOQLHQRLARRNPPOPKLGVA TPISARFQPKLTAVAESVLEGTDTTOSPLKQOS 295
DB 20 SELTRTHKRL-RNRGDPV-----LATLLPATVQANATDEDEE-LDPVVRTAPVLSFHS 71
QY 296 MLKSGS-----AGVTPL-AVTLQK-----KLQAPD-----NPPALNTLLKQ 332
DB 72 DKEGTGEKEVTENWGIYFDNKGVLKAGAILKAGDMLKXKQXTDEXYTNASSFTYSLK 131
QY 333 TLGKDTQYLAHHAHSDGSHLLDNKGHLEFDIKSTATSYSVLHNSHPGEIKGLAQAGT 392
DB 132 DLTDLT-----SVATEKLSFGANGDKVDITSANGL-----KLAKTGN 169
QY 393 GSVSVDS-----KSGKISLGSQTQSH-NKTMLSQPGENHRSLLTGTHQHPAGAR 441
DB 170 GNHLNGLDSTLPDAVNTGLVSSSFTPNDEKTRAA-----TVKDVNLNAGW-NIKGAKT 224
QY 442 PQG--ESIRLHD--DKTHILPELQVMSADKDT-----TGKNTLDVLTAKENKXKTEVKFTPKTSVIK 276
DB 225 AGNVESVDLSAYNNVEFI-----TGKNTLDVLTAKENKXKTEVKFTPKTSVIK 276
QY 478 QADGKLYALKDNRLQNLSDNKSSEKLVDKJKSYSDVORGQVAIITDTPGRHKMSIMPSL 537
DB 277 EKDGKLTGKENNDTNKVTSNATDN-----TD-EGNGLVTA KAVI 316
QY 538 DASPESHLSLHFADAHQGLLHCKSELEAQSVAISHGLRVVADSEGLFSAAIPKQDGG 597
DB 317 DAVNKAQWRVKTITANGONG-----DPATVAGTNTVTFES-DGDTTASVTKDTNG 365
QY 598 NELKWKAMPQHALDEHFGHDHQISGFFHDDHQLNALVKNPNRQOQHACPLGNDHQHPQW 657
DB 366 NGIIVK-----YDAKVGDGLKFPD--- 383
QY 658 NLTDALVIDNOLGLHHTNPPEHIL-----DMGHLSGLALQEGKLHFDQLTKGWIG-A 710
DB 384 --SDKKIVADTTALTVTGKVAETAKEDDKKLVNAGDLVLTALGNL-----SWKAKA 433
QY 711 ESDCKQLKKGLDGAAYLLKDCGPV-----KRLNINQS-----TSSIKHGTENVFSL 755
DB 434 EADTDGALEGIS-----KQDEVKAGETVTFKAGNKLKVKQGCANFTYSLQDALTGLTSI 487
QY 756 PHVRNKPEPGDALGNLKKDRAQAMA-----VIGVNKYIALTEKGDIRSFIKPGTQQLER 811
DB 488 TLGGTTNGNDAKTVINKDGLTITPAGNGTTGTNT-ISVTKDG-----IRAGNKAITN 540
QY 812 PAQTL-----SREGISGELKDI--HVD-----HKQNLALTHTHEGEVFPQ 849
DB 541 VASGLRAYDDANFDVNLNNSATDLNRHVDAYKGLLNLNEKNANKOPLVTDSTAATVGDLR 600

850 REAW-----QNGAESS-----WHKLALPOSESKLSLDM 879
601 KLGWVSTKNGTKESNQVQKQDEVLFTGAGATVTSKSENGKHTITVSVAETK---ADC 657
880 SHEHKPIATFEDGSQHLKAGGWHAYAAPERGLAVGTSQTVFNRLMOGVKGVIPGS 939
658 GLE-----KGDGTIKLV-----DNQNTDNVLTVCNNGTAVTKG 691
940 GL-TVKLSAQTCGMTCAEGRKVSSEKFSERIRAYAFNPTMSTPRPKNAAYATOHGWOGR 998
692 GFETVKT-----GATDADRKVT-----VKDATANDADKKVATVKVATVAINSA----- 735
999 GLKPLYEMOGALIKOLDAHNRHNAPOPDLOS-----KLETLDLGEHGAELNDMKRF 1051
736 ---ATFVKTENLTSSIDEDNPTDNGKDDALKAGDTLTFKAGKNLKVDRDKNIITFDLAKN 792
1052 RDELEQSATRSVTVLGHQGVLSNGEINSEFKPSPOKALVQSFNVRNSGDG----- 1104
793 LEVKTAKVSDTLTIGGN-----TPTGTTATPKVNIITSTADGLNFAKETA 837
1105 ----SKSLOQ---AVHATPPSAESKLSQMLGCHFYVSAGVDSHOKCEPL----- 1146
838 DASGSKNVYLKGIATTLTPEPSAGAK-----SSHVDLNVDATKKSNAASIEDVLRAWNQI 892
1147 --GRORD-----PNDKTALTKSRLIDTVTI-----GELHELADKAKLVSDHK 1187
893 GNGNNVYVATYDTVNFSTDSTGTT-----TVTVTKADGKADYKIGAKTSVIKDH 945
1188 PDADQIKQLRQFED--TLREKRYENPVKHYDMGFTHINKALEANYDAVAFINAFKKEH 1245
946 GKLTCKDLKDANNGATVSEDDGKD-----TGTLVTAKTV---IDAVNK--SGWRVTG 994
1246 HGVLNLTTRTVLESQGSAAELAKLKLNTLLSLDSESMSFSRSYGGVSVFVEPTLSKKVPV 1305
995 EGATAETGATAVAGNAE-----TVTSGTSVNFK---NGNATITATV----- 1032
1306 PVIPIGAGITLDRAYNLSFSRTSGGLNYSFGRDGGVSGNIMVATGHDVMP-----YMTGKK 1360
1033 -----SKDNGNINVKY--DVNVGDLKIGDKDKKIVADTTLTVTGCK 1072
1361 TSAGNASDWLSAKHKISPDLRIGAAVSGTLOGTIONSLLKFLTEDELPGFIHGLTHGLT 1420
1073 VSPAGANSVNNKKL-----VNAEGLATALLNLSWTAKADK---YADGESEGE-- 1118
1421 PAELLQGIHQKQSGKLTFSVDTSANLDRAGINLMDGSK-----PNGVTARVSA 1473
1119 -----TQEVKAGDKVTF-----KAGKNLKVQSEKDFYTSYSLQDTLGLTSI 1160
1474 GLSASANLAAGSRERSTTSGQFSTTSASNRRPTFLNGVGAGANLTAALGAHSHSHEGK 1533
1161 TLGGTAN---GRNDGTGVINKDGLTITLAN-----GAAAGTDSANGNTISVTKDI 1208
1534 PVGIFPAFTSTNVSAAAL-----ALDNRPTSQISLSELEKRAEPV-----TSNDISELTS 1580
1209 SAG---NKEITNVKSALKTYKDTONTADETODKEFHAAVKNAKNEVEFVKNGKATVSAKT 1265
1581 TLGKHFK--DSATTKMLAALKELDDAK-----PAEQHLILQOHFSAKD 1621
1266 NNGKHTVTIDVAEAKVGDLKEDTDGKIKLVNDTNGNLLTVDATKASVAKGEFNA-- 1323
1622 VVGDERYEAVERNK---KLVIROQAADSHSMELGSASHSTTYNNISRNNDGIVELLHKH 1678
1324 VTTDTATTQAQTNANERKGVVVK-----GSGNATATETDKKKVATG----- 1364
1679 FDAALPASSAKRLCEMMNDPALKDI IKQLOSTPFSSASVSVMELKDGLREQTEKAILDCK 1738
1365 -DVAKAIINDAATFYKVENDDSATID-----DSPTDGDANDALKAXDTL---TLKAGKNUK 1415
1739 VGREEVGVLFODRNNLRVKSVSOSVSKSEGFNTPALLLGTSNSAAMSERNIGT INF- 1797
1416 VKRDKGNITFALANDLSVKSATVSDKLS-----LGT-NGKNVNIITSDTKGLNFA 1463
1798 ---KYGQDQNTPRRFTLEGGIAQNPQVASALTD 1828
```

Db 1464 KDSKGTGDAN-----IHLNG-----IASTLTD 1485

RESULT 3

US-08-769-309A-5

; Sequence 5, Application US/08769309A

; Patent No. 5741890

; GENERAL INFORMATION:

; APPLICANT: Scott, John D.,

; APPLICANT: Nauert, Brian J.,

; APPLICANT: Klauack, Theresa M.

; TITLE OF INVENTION: Protein Binding Domains of Gravin

; NUMBER OF SEQUENCES: 24

; CORRESPONDENCE ADDRESS:

; ADDRESS: Marshall, O'Toole, Gerstein, Murray & Borun

; STREET: 6300 Sears Tower/233 South Wacker Drive

; CITY: Chicago

; STATE: Illinois

; COUNTRY: United States of America

; ZIP: 60606-6402

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/769,309A

; FILING DATE:

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: No. 5741890and, Greta E.

; REGISTRATION NUMBER: 35,302

; REFERENCE/DOCKET NUMBER: 27866/33451

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 312-474-6300

; TELEFAX: 312-474-0448

; TELEX: 25-3856

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1780 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-769-309A-5

Query Match 2.2%, Score 203.5; DB 1; Length 1780;

Best Local Similarity 17.8%; Pred. No. 1.9e-05;

Matches 338; Conservative 279; Mismatches 717; Indels 565; Gaps 81;

QY 29 QQSSSSSPQNAASLAAGNKGKMPRIHQPTAAG-----ISAAHQKKSFSIRGCL 83

Db 245 EQSHAEISPPAASGQAVECKEKEGEBKQEPSKSAESPTSPVTSETGSTFKKFTQGWA 304

QY 84 GTKK---FSRSPAGOPCTTHSKGATRLDLLARDGQTHQEAADAAALRTSGGVKRRN 140

Db 305 GWRKTSFRKPEKDEVEASEKKQEPKVDTEEDGK-----AEVASEKLTASEQAHPQE 359

QY 141 MDMAGRPVMVKGSGEDKVPQOQRHQLNFFQMRQTMLSKMAHPASANAGDRLQHSPPH 200

Db 360 PAESAHEPRLSAEYKELPSEEQ----- 383

QY 201 IPGSHHEIKPEVPGSTKATTAHADRVIAEQDDDDSEF-----QQLHQORLARERE--N 252

Db 384 VSGSQGSEEPK---APLATEVFEKIEVHVEVVAHVSVTEERTQETVEETAGS 440

QY 253 PPQPKLGV-ATPISARFQPKLTAVAESVLEGTDTTO-SPLKP-QSMLKGSAGV---TP 306

Db 441 VPABELVGMADPEAPAEPAKELVKLTCVSGEDPTQCADLSPDEKVLKPKPEGVVSEVE 500

QY 307 LAVTLDRKGLQALPDNPPALNTLLKQTLGK-----DTQHYLAHHASSDGSQHLL 355

Db 501 MLSSQERMKVQGSPLKLTSTGLKKLSGKKQKQKRGGGDEESGEHTQVPADSPDQOE-- 558

QY 356 LDNKGHLFDIKSTATSYSLNHPGEGIKGLAQAGTGSVSDGKSGKISLGSQTOSHNK 415

Db 559 -EQKG-----ESSASS-----PEEPFEE--TCLEKGLAEVQDGEAE--GATSGEKK 602

QY 416 TMLSQPGEAHRSLLTGIWQHPAGAARPOGESIRLHDDKTHILHPELVGWSADKDTHSOL 475

Db 603 REGVTPWASFKNVT-----PKRVRPSES-----DREDELKVKASATLSTSTASEM 652

QY 476 SRQADGKLYALKDNRTLNQNSDKSEKLV-----DKIKSYSDVQGOVAI-----L 522

Db 653 QEEMKGSVEEPKPEPKRKVDTSWEALICVGSKKRARRRRSSDEEGCPKAMGDHOK 712

QY 523 TDTGRHKMSIMPSILDASPEHSLSLHFADAHQ-----LLHGKSELAQSAVISHGRL 577

Db 713 ADEAGKDKETDGLAGSQEH-----DPGQSSSPPEQAGSPTEGEGVSTWESFKRL 764

QY 578 VVADSEGLFSAAIIPKQGDGNEKMKAMPQHALDEHFGHDHQSIFGFFHDDHGLNALVKN 637

Db 765 VT-----PRKKSCKLEBS-----EDSIAG----- 785

QY 638 NFRQOHACPLGNHQFHPGNLTDALVTDNQLGLHHTNPHEPHEILDHGLSLAQEGKL 697

Db 786 -----SGVHSTPDE----- 796

QY 698 HYFDLTKGTGTAESDCKQLKKGLDGAAYLLKDGVEKRLINQSTSSIKHGTVNFSLPH 757

Db 797 -----PKESWSVSIKFFIPGRKKRPDQKQAPVEDAGPT----- 833

QY 758 VRNKPEPDALQGLNKDDKAQAMVIGVNYKYLALTEKGDIRSFIKPGTQQLERPAQTLS 817

Db 834 -----GANEDD-SDVPAVPLSEYDAV-EREKMEAAQQAQKGAEPQKAAT-- 877

QY 818 REGISGELKDIHVDHKQNYLA-LTHEGVFHPQPRAWONGABSSSHKL-----ALPQSESK 873

Db 878 --EVSKELSQVHMMAAAVADGTRAATIIERSPSWISASVSTPELEQVAAALITTEV 935

QY 874 LKSLDMSHEHKPIATFEDGSGHQLKAGGHAYAAAPERGLAVGTSGSTVFNRMLQGVKG 933

Db 936 LEREVIAEEEPPTV-----EPLPE-----NREARG--D 962

QY 934 KVIPGSLTVK--LSAQGTGTMGA-EGRKVSFKFSERIRAYAFNPTMTSPRIKNAAYAT 990

Db 963 TVVSEAEITPEAVTAETAGPLGSEEGTEASAEETEMWSAVSQTLDSPDTEBAT-- 1019

QY 991 OHGQMGREGKPLYEMOGAL-----IKQLDAHNRHNAPOPDLOSKLETLDLG 1038

Db 1020 -----PVOEEGGVPDIEEQERRTOEVLQVAEKVKEESQLP-----G 1057

QY 1039 EHGAELLDNMRFRDELEQASRSTVVLGQHQGVLSNGEINSEFKPSP---GKALVQS 1094

Db 1058 TGGPEDVLQPVQRAEAERPEQEAASGLKKTETDVVLKDAQ---EAKTEPTQGVV--- 1111

QY 1095 FVNRSGDLSLQQAHVATPPSAESKLSMLGHFVSAGVDMHQKGEIPLGRQDRND 1154

Db 1112 -----GQTTPESEKAPQVTESTESSELVTTCAETLAGV-----KSEQVMVEQAIPD 1160

QY 1155 KTALTKSRLLDVTIGELHELADKAKLVSDHKPDADQIKOLRQOQDFLREKRYESNPVK 1214

Db 1161 S-----VETPTSETGSPVADFDAPGTTQKQKEIVEI-----HEENEV- 1199

QY 1215 HYTDMGFTHNKALEANYD---AVKAFINAFKKEHHGNLTATRTVLESQGSALAKKNT 1271

Db 1200 HLPVVRGTEAEAVPAQKERPPAPSSFV--FQEE-----TKEQSKMEDT 1240

QY 1272 LLSLD---SGESMS-FSRSYGGGVSTVFTLSKKVPVPIPGAGITLDRAINLSFSRTS 1327

Db 1241 LEHTDREVSVEVTSILSKTEGTQADQADEKTKDVPF-----FE--- 1280

QY 1328 GGLNVSFGRGGVSGNIMVATGHVMPYMTGKTKSAGNA--SDWLSAK-HKISPDRLIGA 1384

Db 1281 -----GLEGSIDTGTIVSREKVEVALAGEGTEAECKDDALELQSHAKSP----- 1327

QY 1385 AVSGTLOGLTQNSLKFXTLDEBLPGFIHGLTHGLTTPAELLQKGIEHOMKOGSKLTFSD 1444  
Db 1328 --PSPVEREMVQVVEREKTEAE-----PTHVNEEKLEHE---TAVTVSEE 1367  
QY 1445 TSANLDRACINLNE-----DGSKP-----NGVTARVSAGLS-ASANLAAGSRERS 1489  
Db 1368 VSKOLLTVNVPIDGAKAEYSSLEGGPPCLGQEEAVCTKIQVOSSEASFLLTAAAEK 1427  
QY 1490 TTSGQGSTTSASNRRPFLGVGAGANLTAALGVAHS-----STHEGK---PVG----- 1536  
Db 1428 V----LCETANILETGETL---EPAGAHVLEEKSEKNEKEDFAAHPGEDAVPTGPDCAK 1480  
QY 1537 ----IPAFSTNVSAALADNRFS-----QSTSLKRAEPVTSNDIS 1576  
Db 1481 STPVIVSATTKGLSSDLEGETTKTLKWKSDVDEQVACQEVKVSVAIEDLEP---ENGIL 1538  
QY 1577 ELASTLKGKHPK-----DSATTKMLAALKELDDAKPAEQHLIL-----QQHPSA 1619  
Db 1539 ELETSSKLVONIQTAVDQVFRTEETATEMLTS--ELQ-----TQAHVIKASODAQOE 1591  
QY 1620 KVVGVGERYBAVRNLKLVIRQQAADSHMELGSAHSHTYNNISRIINDDGIVELLHKHF 1679  
Db 1592 TEKEGEPOASAO--ETPITSKEESESTAVGOA-HSDISKMS-----EASEKTM 1640  
QY 1680 DAALPASSAKRLGEMMNDPALKDIIKLOLSTPSSASVSMELKDG---LREQTEKALD 1736  
Db 1641 TVEGEGTV-----NDQOLEEVLPSEEGGAGTKSVDPDGHALLAERIEKSILVE 1692  
QY 1737 GYGVREGEVGLFODRNNLRKVSVSQSVSK-SEGFNTP 1774  
Db 1693 PKEDEKDDDDPENQNSALADTDASGLTKPESPTNGP 1731

## RESULT 4

US-08-994-570-5  
; Sequence 5, Application US/08994570  
; Patent No. 6090929  
; GENERAL INFORMATION:  
; APPLICANT: Scott, John D.,  
; APPLICANT: Nauert, Brian J.,  
; APPLICANT: Klauck, Theresa M.  
; TITLE OF INVENTION: Protein Binding Domains of Gravin  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 6300 Sears Tower/233 South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States of America  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/994,570  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: NO. 6090929and, Greta E.  
; REGISTRATION NUMBER: 35,302  
; REFERENCE/DOCKET NUMBER: 27866/33451  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-474-6300  
; TELEFAX: 312-474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1780 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear

## ; MOLECULE TYPE: protein

US-08-994-570-5

Query Match 2.2%; Score 203.5; DB 3; Length 1780;  
Best Local Similarity 17.8%; Pred. No. 1.9e-05;  
Matches 338; Conservative 279; Mismatches 717; Indels 565; Gaps 81;

QY 29 QGSSSSSSPONAASLAAGKNGKMPRIHOPSTAADG-----ISAHHQKKSFSLRCL 83  
Db 245 EQSHAEISPPAESQAVECKEKEGKOEKPSKSAESPTSPVTSEGTGTFKFFTGWA 304  
QY 84 GTRK---FSRSPQPGGTHSGATLRLDLDLARDGGTQHEAAAPDAARLFRSGVKRN 140  
Db 305 GWRKKTSPKPKDEVEASEKKKEQPEKVDTEEDGK-----AEVASEKLTASQAHPOE 359  
QY 141 MDMAGRPVMVGGSGEDKVPTOOKRHQLNFGOMRQMTLSKMAHPASANAGRLQHSPPH 200  
Db 360 PAESAHEPRLSAEYKVELPSEQ----- 383  
QY 201 IPGSHHEIKKEPVGSTSKATTAHADRVETIAQEDDDSEF-----QQLHQOBLARERE--N 252  
Db 384 VSGQSPSEKP--APLATEVFEDEKIEVHQEEYVAEVHVSVEERTEEQKTEVEETAGS 440  
QY 253 PPQPPKLGV-ATPISARFOPKLTAVAESVLEGTDTTO-SPLKP-QSMLKGSAGV---TP 306  
Db 441 VPAEELVGMDAEPORAPAKELVKLCTCVSGEDPTOGADLSPEKVLSPKPGVSEVE 500  
QY 307 LAVTLDKGLQLADNPPLNTLLKQTLGK-----DTQHYLAHHASSDGSQHL 355  
Db 501 MLSSQERMVQGSPLKLLFTSTGLKLSGKKQKGRGGDEESGEHTQVPADSPDSQE-- 558  
QY 356 LDKGHGLFDIKSTATSYSVLHNSHPGKIKLAQAGTGSYSDGKSGKISGSGTQSHNK 415  
Db 559 -EOKG-----ESSASS-----PEEPEI--TCLEKGLAEVQOQGEAE--GATSDGKK 602  
QY 416 TMLSQPGEAHRSLLTGIWQHPAGAARPGESIRLHDDKIHLHPGLGVWQASADKDTSQL 475  
Db 603 REGVTPWASFKKMVT-----PKKRVRRPSES-----DKEDELKVKSATLSTESTASEM 652  
QY 476 SRQADGKLYALKDNRTLQNLSDNKSSEKLV-----DKIKSYSDVORGVAI-----L 522  
Db 653 QEEMKGSVEEPKPEPKRYDTSVSWBALICVSGSKRRARRSSDEEGGPKMGMDHQK 712  
QY 523 TDTPPGRHKMSIMPLSDASPESHISLSLHFADAHG-----LLHGKSELEASVAISHGRL 577  
Db 713 ADEAGKDKETGDTGILAGSQEH-----DPGGSSSPQOAGSPTEGEGVSTWESFRL 764  
QY 578 VVADSEGLFSAAIPKOGDGNELKMKAMPQHALDEHFCHDHQISGFFHDDHGLNALYKN 637  
Db 765 VT-----PRKSKSKLEKS-----EDSIAG----- 785  
QY 638 NFRQOACPLGNDHQHPGWNLTALVIDNLGLUHTNPPEPHEILDMGLSLALQEGKL 697  
Db 786 -----SGVEHSTPDT----- 796  
QY 698 HYFDQLTKGWTGAESDCQKLGDAAYLLKDCGEVRLNINQSTSSIKHGFENVESLPH 757  
Db 797 -----PGKEESWVSISKPIGRRKKRPDKQEQAPVEDAGPT----- 833  
QY 758 VRNKPEPDALQGLINKDDKAQAMAVIGVKNYKLYALTERGDIRSFQIKPTQOQLERPAOTLS 817  
Db 834 -----CANEDD-SDVPVAVPLSEYDAV-EREKMEAAQQAQKAEQPEKKAAT-- 877  
QY 818 REGISGELKDIHVDHKONLYA-LTHEGEVPHQPREANONGAESSWHLK---ALPOSESK 873  
Db 878 --EVSKELSQVHMMAAAVADGTRAATIIIEERSPSWISASVTEPLEQVEAEALLTEEV 935  
QY 874 LKSLDMSHHKPIATPTEDEGSOHLKAGWHAYAAPERGPLAVGTSGSQTVFNRLMOGVKG 933  
Db 936 LEREVIAEEEPPTVT-----EPLPE-----NRARG--D 962  
QY 934 KVIPGSGLTVK--LSAOTGGMGTGA-EGRKVSSKFSERIRAYAFNPTMTSTPRPIKNAAYAT 990

```
Db 963 TVVSEALTPPAVTRAEATAGPLGSEGTAEASAAETTEMVSAVQSLTDSPTTEAT--- 1019
QY 991 QHGWQGRGLKPLYEMOGAL-----IKOLDAHNVHRNAPQPDQLSKLETLDLG 1038
Db 1020 -----PVQVEGVGPDIEQERTQEVLOQAVEKVEESQLP-----G 1057
QY 1039 BHGAE-LLNDMKRFRDELEQSNATRSVVLQGHGVLSKNGEINSEKPKSP---GKALVQS 1094
Db 1058 TGGPEDVLQPVQRAERPEQAEASGLKKTEDVVLQVDAQ---EAKTEPFTQGVV--- 1111
QY 1095 FNVNRSGDLSKSLQOAVHAPPSAFESKLSQMLGHFVSAGVDMSHQGEIPLGRQDPND 1154
Db 1112 -----GQTTPESEFKAPQVETIESSELVTTTCAETLAGV-----KSOEMVMEQAIPPD 1160
QY 1155 KTALTKSLIILDTVTIGELHELADAKLVSDHKPADQIKQLRQFDTLREKRYESNPVK 1214
Db 1161 S-----VETPTDSETDSTPVADPDAPGTTQKDEIVEI-----HEENEV- 1199
QY 1215 HYDMGFTHNKALEANYD---AVKAFINAFKKEHHGVNLTTRTVLESQGSAAELAKLNT 1271
Db 1200 HLVPVRGTAEAVPAOKERPAPSPFV--FQEE-----TKEQSKMEDT 1240
QY 1272 LLSLD---SGESMS--FRSYGGVSTVPVPTLSKKVPVPIPCAGITLDRAYNLSFRSTS 1327
Db 1241 LEHTDKEVSVETVSLSKTEGTQEQADQADEKTKDVPF-----FE--- 1280
QY 1328 GGLNVSGRQDVGSGNIMVATGHDVMPYMTGKTKSAGNA--SDWLSAK-HKISPDRLRIGA 1384
Db 1281 -----GLESGIDTGVTSREKYTEVALKGEGTEAECKDDALEQSHAKSP----- 1327
QY 1385 AVSTLOGLTQNSLKFUKTEDELPGFIHGLTHGTTTPAELLQKIBHQMKQSGKLFTSVD 1444
Db 1328 --PSPVEREMVQVVEREKTEAE-----PTVNEEKLHE-----TAVTVSEE 1367
QY 1445 TSANLDRAGINLNE-----DGSKP-----NCVTRVRSAGLS-ASANLAAGSRERS 1489
Db 1368 VSKLLQTVNVPVITIDGAKEVSLSGPPPCIGQEERAVCTKIQVOSSEASFTLTAAAEK 1427
QY 1490 TTSQFGSTTSASNNRPTFLNGVAGANLTAALGVAHS-----STHEGK---PVG----- 1536
Db 1428 V---LGETANILETGETL---EPAGAHVLEEKSEKNEKEDFAAHPGEDAVPTGPDQCAK 1480
QY 1537 -----IFPAFTSTNVSAAALNDRTS-----OSISLELKRAPVTSNDIS 1576
Db 1481 STPVIVSATTKKGLSSDLEGEKTTLSKWKSDVDEQVACQEVKVSVAIEDLEP--ENGIL 1538
QY 1577 ELTSTLGHKHF-----DSATTKMLAALKELDDAKPAEOLHIL---QGHFSA 1619
Db 1539 ELETSSKLVQNIITQAVDQFVRIETETATEMLTS--ELQ-----TQAHVIKADSQAGQE 1591
QY 1620 KDVVGDERYEAVRNKLVIRQQAADSHSMELGSAHSTTYNNLSRINNDGIVELLHKHF 1679
Db 1592 TEKEGEPQASQAD--ETPITSAKESESTAVGQA-HSDISKDMS-----EASEKTM 1640
QY 1680 DAALPASSAKRLGEMMNDPALKIILQLOSTPFPSSASVSMELKDG---LREOTEKAILD 1736
Db 1641 TVEEGSTV-----NDQOLEEVLPSEEGGAGTKSVPEDDGCHALLAERIEKSLVE 1692
QY 1737 GKVGREEVGLVFODRNLRVKSVSQSVSK-KSEGFNTP 1774
Db 1693 PKDEKGGDDVDDPENONSALADTDASGGLTKRESPDTNGP 1731
```

## RESULT 5

```
US-09-377-155-33
; Sequence 33, Application US/09377155
; Patent No. 6197312
; GENERAL INFORMATION:
; APPLICANT: PEAK, Ian Richard Anselm
; APPLICANT: JENNINGS, Michael Paul
; APPLICANT: MOXON, E. Richard
; TITLE OF INVENTION: NOVEL SURFACE ANTIGEN
```

```
FILE REFERENCE: 065064/0128
CURRENT APPLICATION NUMBER: US/09/377,155
CURRENT FILING DATE: 1999-08-19
PRIOR APPLICATION NUMBER: PCT/AU98/01031
PRIOR FILING DATE: 1998-12-14
PRIOR APPLICATION NUMBER: GB 9726398.2
PRIOR FILING DATE: 1997-12-12
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 33
LENGTH: 2353
TYPE: PRT
ORGANISM: Haemophilus influenzae
US-09-377-155-33
```

```
Query Match 2.1%; Score 201; DB 4; Length 2353;
Best Local Similarity 18.9%; Pred. No. 4.5e-05;
Matches 346; Conservative 223; Mismatches 665; Indels 592; Gaps 84;
```

```
QY 236 SEFOQLHQORLARERENPPQPKLGVATPISARFQPKLTAAVESVLEGTDTTQSPKPKQS 295
Db 20 SELTRTHTRKASATVETAV-----LATLLFAIVQANATDEDEE-LDPVVRTAPVLSFHS 72
QY 296 MLKGGG-----AGVTPL-AVTLDDKGLQALAPDNPALNTLLKOTPLGKDT 338
Db 73 DKEGTCEKEVTENSNWGIYFDNKGVLKAGAITLKAGD-----NLKIKQNTDEST 121
QY 339 QHYLAHHAASSQSHLLDNKNGHLFDIKSTAT-SYSVLHNSHPGEIKG-----KLAQAGT 392
Db 122 -----NASS-----FTYSLKDLTLTSVATEKLSFGANGDKVDITSDANGLKLAGTGN 170
QY 393 GSVSYVDG-----KSGKISLGSQTQSH-NKTMLSQPEAHRSILITGIWOHPAGAAR 441
Db 171 GNVHLNGLDSTLTPDAVTNGVLSSTSPNDVEKTRA-----TVKDVNLNAGW-NIKGAKT 225
QY 442 PQG--ESTRL-----HDDK-----IHLPELGVQWQSAKDTHLSQSRQADGKLYA 485
Db 226 AGGNVESVDLVAYSANNVEFITGDKNTLDVVLTAKENGKTEVKFTPKTSVKEKDGKLT 285
QY 486 LKDNRTLQNLNPKSSEKLVDKIKSYSDQGVQVAILTDPGRHKMSIMPSLDASPESHI 545
Db 286 KENNDTNKVTSTATDN-----TD-EGNGLVTAKAVIDAVNKAAGW 325
QY 546 SLSLHFADAHQGLLHGKSELAQSAVSHGRVYVADSEGLFSAAIPKQDGNELMKAM 605
Db 326 RVKTTTANGQNG-----DFATVASGTNTVTFES-GDGTASVTAKDTNGNGITVK-- 372
QY 606 PQHALDEHGHQHSQFFHDDHGLNQLNALVKNFRQOHACPLGNDHOFHPGNLTLDAVI 665
Db 373 -----YDAKVGDGLKPD-----SDKKIV 390
QY 666 DNQLGLHHTNPPEHIL-----DMGHLGSLALQEGKLYHFDQLTKGTWG-AESDCKOLK 718
Db 391 ADTTALTVTGKVAEIAEKEDDKKLVNAGDLVLTALGNL-----SWKAKAEADTGAL 442
QY 719 KGLDGAAYLLKQGEV-----KRLNINOS-----TSSIKHGTENVFSLPHVRNKP 763
Db 443 EGIS-----KDQEVKAGETVTFKAGKNLKVQDQGANFTYSLQDALTLTSLTGGTTNG 496
QY 764 PGDALOGLNKDDKAQAMA-----VIGVNYKYLALTEKDIRSFQIKPGTQOOLERPATL--- 816
Db 497 GNDAKTVINKDGLTTPAGNGGTTGTNT-ISTVKDG-----IKAGNKAITNVASGLRAY 549
QY 817 ---SREGISGELKDI-HVD-----HKQNLVALTHEGEVVFHOPREAW----- 853
Db 550 DDANFDVLNNSATDLNRHVEDAYKLLNLNEKNANKOPLVTDSTAATVGDRLKLGWVST 609
QY 854 QNGAESSE-----WKKLALPQSESKLSKSLDMSHEHKPIA 887
Db 610 KNGTKEESNOVKQADEVLFTGAGAAATVTSKSENGKHTITVSAETK---ADCGLE----- 661
QY 888 TFEQSSQHLKAGGWHAYNAAPERGPLAVGTSGSOTVFNRLMQGVKGVIPCSGL-TYKLS 946
```



Db 286 GKENDDTKNTSNTADN-----TD-EGNGLVTAKAVIDAVNKGAGW 325  
QY 546 SLSHFADAHOGLLHGKSELEAQSVAISHGRVLVWADSEGLKPSAAIPKQGDGNELKMKAM 605  
Db 326 RVKTTTANGONG-----DFATVASGTVTFES-GDGTASVTKTOTNGNGITVK-- 372  
QY 606 PQHALDEHFGHDHQISGFFHDDHGGQNALVKNFRQOHACPLGNDHQFHPGNLTALVI 665  
Db 373 -----YAKVGDDGLKFD-----SKKIIV 390  
QY 666 DNQGLHHTNPPEHIL-----DMGHGLSLALQEGKLFYFDQLTGKWTG-AESDCKQLK 718  
Db 391 ADTALTVTGKVAEIAEKDDKKLVNAGDLVTALGNL-----SWRAKAEADTDGAL 442  
QY 719 KGLDGAAYLLKDGVE-----KRLNINQS-----TSSIKHGTENVPSPHVRNKPE 763  
Db 443 EGIS-----KDQEVKAGETVTFKAGKNLKVKQDGFANFTYSQDALTGTLTSLTGFTNG 496  
QY 764 PGDALQGLNKDDKAQAMA-----VIGVNYKYLALTEKDIRSFQIKPGTQOLERPAQTL--- 816  
Db 497 GNDAKTVINKDGLTTPAGNGTGTGNT-ISTVKDG-----IKAGNKAITNVASGLRAY 549  
QY 817 ---SREGISGELKDI--HVD-----HKONYALTHEGEVHFHQPREAW----- 853  
Db 550 DDANFDVLNNSATDLNRHVEDAYKGLLNLNEKNANKQPLVTDSTAATVGDRLKLGWVST 609  
QY 854 ONGAESSE-----WHKLALPOSESKLSLSDMSHEHKPIA 887  
Db 610 KNGTKEESNQVKADEVLTFTGAGATVTSKSENGKRITITVVAETK---ADCGLE----- 661  
QY 888 TFEQSOHQKAGGWAYAAAPERGVLAVGTSGTSQTVFNRLMOGVKGVIPGSGL-TVKLS 946  
Db 662 --KQSDTTLKLV-----DNQNTDNLVTGNGTAVTKGGFEIVKT- 699  
QY 947 AQTGCMGAEGRKVSSEKSERTRAVAFNPTMTSTPRIPKNAAYATQGWQREGKLPDIEM 1006  
Db 700 ---GATDADRGRV-----VKDATANDADKKVATVKDVATAINSA-----ATFVK 741  
QY 1007 QGALIKOLDANVRHNAPOPDLOS-----KLETLDLGEHGAELNDMKRFEDELQSA 1059  
Db 742 TENLTSSIDENPTDNGKDDALKAGDTLTFKAGKNLKVRDGNKIITFDLAKNLEVKTAKV 801  
QY 1060 TRSVTVLQHOGLVKSNGEINSEFPKSPGKALVQSPNVNRSGDL-----SKSL 1108  
Db 802 SDTLTIGGN-----TPTGGTTATPKVNIITSTADGLNFAKETADASGSKNV 846  
QY 1109 QQ---AVHATPPSAESKLSQMLGHFVSAGVDMSHQKEIPL-----GRQD- 1151  
Db 847 YLKGITATLTPEPSAGAK-----SSHVDLVNADTKKNAASIEDVLRAWNIOGNNVNDY 901  
QY 1152 -----PNDKTALTKSRLLDVTI-----GELHELADKALVSDHKFPDADQIKQ 1195  
Db 902 VATYDVTVNFTDSTGTT-----IVTVQKADGKADVKIGAKTSVIKDHNGKLFQTKD 954  
QY 1196 LROQPD--TLREKRYESPVKHYITDMGFTTHNKALEANYDAVKAFIAPKKEHGVNLTTR 1253  
Db 955 LKDNANGATVSDDDGD-----TGTGLVTAKTV-----IDAVNK--SGWRVTGEGATAETG 1003  
QY 1254 TVLESOGSAELAKLKNLTLLSDSGESMSFSRSGGVSTVFVPTLSKKVPVPVPCAGI 1313  
Db 1004 ATAVNAGNAE-----TVTSGSVNFK-----NGNATTATV----- 1033  
QY 1314 TLDRAYNLSFRTSGGLNVSGFRDGGVSGNIMVATGHDVMP-----YMTGKKTSGNASD 1368  
Db 1034 -----SKDNGNINVKY--DYNVGDGLKIGDDKKIVADTTTLTVTGKVSVPAGAN 1081  
QY 1369 WLSAKHKTPDLRIGAAVSGTLOGLTQNSLKFKLTEDELPGFIHGLTHGTLTPAELLQK 1428  
Db 1082 SVNNKKL-----VNAEGLATALNNLSWTAKADK---YADGESEGE----- 1119  
QY 1429 IEHOMKOGSKLTFSDTANLDRAGINLEDGSK-----PNGVTARYSAGLSASANL 1481  
Db 1120 TDQEVKAGDKVTF-----KAGKNUKVKQSEKDFYSLQDTLTGLTSLTILGGTAN- 1168

QY 1482 AGRSRSTTSQFGSTTSASNNRPTFLNGVGAGANLTAALGVAHSSTHEGKPVGIFPAF 1541  
Db 1169 --GRNDTGVINKDGLTITLAN-----GAAAGTASNGNTISVTKDGLISAG---NK 1214  
QY 1542 TSTNYSAAAL-----ALDNRTSQSISLELKRAEPV-----TSDNISLSTLGLKHFK- 1587  
Db 1215 EITNVKSALTKYKTONTADETKQKEFAAANKVANEVEFFVGKNGATVSAKTDNNGKHVT 1274  
QY 1588 -DSATTKMLAALKELDDAK-----PAEQULHILOQHSKADVVGDERYE 1629  
Db 1275 IDVAEAKVGDLGKEDTGKIKLVQDNTGNLLTVDATKASVAKGEFNA--VTTDATTA 1332  
QY 1630 AVRNELK---KLVIROQAADSHSMELGSASHSTVNNLSRINNDGIVELLHKKHFDALPAS 1686  
Db 1333 QGTNANERKVVVK-----GSGNATATETDKKKVATVG-----DVAKAIN 1372  
QY 1687 SAKRLEGMMNDPALKDIIKQLQSTPSSASVSMELDGLREOTEKAILDKGVGREGVGV 1746  
Db 1373 DAATFVKVENDSATID-----DSPDTDDGANDALKAGDTL---TLKAGKNLKVRKDKNI 1424  
QY 1747 LFQDRNNLRVKSVSQSVSKSEGFNTPALLGLTSSAAMSERNIGTINF-----KYGOD 1802  
Db 1425 TFLANDLSVKSATVSDKLS-----LGT-NGNKVNITSDTKGLNFAKDSKTGDD 1472  
QY 1803 QNTPRRTLEGGAQANPOVASALTD 1828  
Db 1473 AN-----IHLNG-----IASLTLD 1486

RESULT 7  
US-08-038-682-2  
; Sequence 2, Application US/08038682  
; Patent No. 5549897  
; GENERAL INFORMATION:  
; APPLICANT: BARENKAMP, STEPHEN J  
; APPLICANT: ST. GEME III, JOSEPH W  
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS  
; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Shoemaker and Mattare, Ltd  
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202-0286  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/038,682  
; FILING DATE: 16-MAR-1993  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BERKSTRESSER, JERRY W  
; REGISTRATION NUMBER: 22,651  
; REFERENCE/DOCKET NUMBER: 1038-293  
; TELEPHONE: (703) 415-0810  
; TELEFAX: (703) 415-0813  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1536 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-038-682-2





ATTORNEY/AGENT INFORMATION:  
NAME: Berkstresser, Jerry W  
REGISTRATION NUMBER: 22,651  
REFERENCE/DOCKET NUMBER: 1038-404  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0813  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1536 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-302-832-2

Query Match 2.0%; Score 190.5; DB 1; Length 1536;  
Best Local Similarity 18.6%; Pred. No. 0.00013;  
Matches 297; Conservative 221; Mismatches 559; Indels 521; Gaps 72;

QY 304 VTPLAVTLDK-----GKQLAPDNPALNTLLKQTLGKDTQHYLAH---HASSD 349  
DB 149 INPNIITIGDAIINTNGFTASTLDISENIKARNFTFEQTKDALKALAEIVNHLITVGD 208  
QY 350 GSOHLL---LDNKG-----HLFDIKSTATSYSVLHNSHPGEIKGLAQ 389  
DB 209 GSVNLIGKVKNEGVISYNGSGSISLLAQKITDISDIINFTIYISAAPENAVNLGDI 268  
QY 390 AGTGSVVDG---KSGKISLGST---OSHNTMLSQPGEAHRSLTGTWHPAGARPQ 443  
DB 269 KG-GNINVRATIRNQKLSADSVSKSGNIVLSAKEGA----- 308  
QY 444 GESIRLHDDKIHLPELVGWSADKTHSOLSROADGLYALKDNRTLQ-----NLSDN 498  
DB 309 -----ELGGVISA-----NQQAQKGGKLMITGDKVTLKGTGAVIDLSGK 346  
QY 499 KSEKLVVDKISYVDQGVAILTDTTPRHKMSIMPSIDASPESHISLHFADAHQGL 558  
DB 347 EGGETYL---GGDERGE-----GKNGIQAKKTSLEKSGTINVS----- 382  
QY 559 LHGSELEAQSVAISHGLRVADSEGLFSAAPKPGDCNELKMKAMPQHALDEHFGHDH 618  
DB 383 --GK---EKGRAIWMGDIALDGN-----INAGSGDIAKTGGFV-----ETSGHD- 424  
QY 619 QISGFFHDDHQLNALVKNFRQOHACPLGNDHQHPGWNLT-DALVIDNOLGLHHTNPE 677  
DB 425 ---LFIKDNAIVDA-----KEWLLDFDNVSINAETAGRSNTSE 459  
QY 678 PHEILDMHGLSLALQEGKLHYFDOLTGWTGAESDCKOLKGLDGAAYLLKDGVEKRLN 737  
DB 460 DDEVTGSGNSASTPKRNK-----EKTTLTNTTLES-----ILKKGTFVNT 500  
QY 738 INQ-----STSIKIGTENVSFL-----PHVRNRPEDGALQGLNKKDKKAQAMAVIGVN 786  
DB 501 ANQRIYVSSNLSNLSNGSLTSEGRSGGGVEINNDITTDGDTREGANL--TIYSGGWVDVH 558  
QY 787 KYALTEKGDGDI-----RSFOIKPGTQOLRPAOTLSREGISGELKDIHVDH----- 832  
DB 559 KNISLGAOGNINITAKQDIAFEKGSNOVITGQTTI---SGNQKGRFPNNVSLNGTSG 614  
QY 833 -----KONLYALTHEGEVFFHOPREAWQNGAESSSWH---KLALPOSES----- 872  
DB 615 LQFTTKRTNKVATNKFE-----GTLNISGVNISMVLPKNESGYDKFKGRTYW 663  
QY 873 KLKSLDMSHEHKPIATFEDGSOHLKAGGWHAYAAPERPLAYGT-----SGSQ 921  
DB 664 NLTSNVSSEGEFNITID-----SRGSDSAGTLTQPYNLNLSNFKD 705  
QY 922 TVFN-----RLMOGVKGVIPGSLTVKLSAOTGGMTGAEGRKVSFKFSRIRAYAFNPT 976  
DB 706 TTFVNRNARVNFIDKAPIGINKYSLNYSFNGNISVSGGSGVD-----FTLLASSN 759

QY 977 MSTPRPIKNAAY-----ATQHGQWQREGKLPXYEMOQALIKQLDAHNVRH 1021  
DB 760 VQTPGVINSKYFNVSTGSSLRFKTSKSTGTGFSIEKDL--TLNATGNITLLQVEGTDG 817  
QY 1022 NAPIQDLOSKELETDLGEGHGAELLNDMKRFRDELEOSAT---RSVTVLG---OHQ--- 1070  
DB 818 MIGKIVAKKNITFE---GGNITFGSRKAVTEIEGNVTINNANVTLLIGSDFDNHOKPL 873  
QY 1071 -----GVLKSNGEI-----NSEFKPSGKALVQSPFNVNRSQ----- 1102  
DB 874 TIKKDVIIINSGLNTAGGNIVNIAGNLTVESNANFK-----AINTFFNVGGLFDNKGNS 927  
QY 1103 -----DLSKSLQQAIVA-----TPSAESKLSQML 1127  
DB 928 NISIAKGGARPKDIDNSKNLSITTSSTVYRTIISGNITNKGDLNITNEGSDTEMQ--- 984  
QY 1128 GHFVSAGVDMSHQGEIPL-----GRQDPNDKTALKSRILDTVTIGE 1172  
DB 985 -----IGDVSQKEGNLTISSDKINIKQITIKAGVDGSDSDATNANLITIKTELK 1039  
QY 1173 LHEL-----ADKAKLVSDHDKPDADQIKQLRQOQFDTLREKRYESNPVKHYTDMGFTHNKALE 1228  
DB 1040 TQDLNISGFNAEITA--KGSDLTIGTNTNSADGTNAKVTFNQVKDSKISADGHKVTLH 1097  
QY 1229 ANYDAVKAFINAFKKEHGVNLTTRTVLESQGSABELAKLKNLTL--SLDSCGSMSPRS 1286  
DB 1098 S-----KVETSGSNNTED--SSDNNAGLTIDAKNVTNNNITSHKAVSISAT 1143  
QY 1287 YGGGVSTVFVPTLSK---KVPVPVIPA--GITLDRAYNLFSFRTSGGLNVSFGRDGGVS 1341  
DB 1144 -SGEITTKGTIINATGNVEITAGTGSILGGIESSGSVTLTATEGALAVS-----NIS 1197  
QY 1342 GNIMVATGHVMPYMTGKTSAGNASDMLSAKHKISPLDRIGAAVSG-----TLQGT 1394  
DB 1198 GNTVTVTANGSALTTLAGSTIKGTSVTTSSQ---SGD--IGGTISGTVVEVKATESLT 1252  
QY 1395 QNSLKFLEDE---LPGFTHLTHGLTLPALLOKQIEHOMKQSGKLTFSVDTSANL 1449  
DB 1253 QSNKIKATGGEANTVSATGTIGGTISGN-----TVNVTANA 1289  
QY 1450 -DLRAGINLNEGSGKPNVGTARYSAG---LSASANLAAGSRERSTTSQFGSTTSASNRR 1505  
DB 1290 GDLTVG-NGAEINATGEAATLTSSGKLITTEASSHITS-AKGQVNLQAQDGVAGSINAA 1347  
QY 1506 PTFELGVGA---GANLTAALGV-----AHSSTHEGKPVGIFPAFTSTNVSAALADNRT 1556  
DB 1348 NVTLNNTGTLTVKGSNINATSGTLVINAKDAELNGAALGNHTVYNATNANGSGSVIAT 1407  
QY 1557 SOSISLELKRAEPTVNDISELTSTLTKKHKFDSATTKML-----AALKELD 1602  
DB 1408 SSRVNI---TGDITITINGLNIIS-----KNGINTVLLKGVKIDVKYIQPGIASVDEVI 1457  
QY 1603 DAKPAEQHLIQHFSAKDVGVGDYEAVERNKLKLVIR 1640  
DB 1458 EAK-----RILEK---VKD-LSDEEREAALAKLGVSAVR 1486

RESULT 9  
US-08-530-198-2  
; Sequence 2, Application US/08530198  
; Patent No. 5869065  
; GENERAL INFORMATION:  
; APPLICANT: BARENKAMP, STEPHEN J  
; APPLICANT: ST. GEME III, JOSEPH W  
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS  
; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Shoemaker and Mattare, Ltd  
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
; STREET: Bldg. 1  
; CITY: Arlington  
; STATE: Virginia



APPLICANT: Barenkamp, Stephen J.  
TITLE OF INVENTION: High Molecular Weight Surface Proteins  
TITLE OF INVENTION: of No. 5876733-Typeable Haemophilus

NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Shoemaker and Mattare, Ltd.

STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202-0286

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/469,880

FILING DATE: 06-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: GB 9205704.1

FILING DATE: 16-MAR-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US PCT/US93/02166

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/302,832

FILING DATE: 16-SEP-1994

ATTORNEY/AGENT INFORMATION:

NAME: Berkstresser, Jerry W

REGISTRATION NUMBER: 22,651

REFERENCE/DOCKET NUMBER: 1038-516 MIS:vg

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 415-0810

TELEFAX: (703) 415-0813

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1536 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-469-880-2

Query Match 2.0%; Score 190.5; DB 2; Length 1536;  
Best Local Similarity 18.6%; Pred. No. 0.00013;  
Matches 297; Conservative 221; Mismatches 559; Indels 521; Gaps 72;

```
QY 304 VTPAVTLQK-----GKQLAPDNPALNTLLKQTLGKDTQHYLAH---HASSD 349
DB 149 INPNTIGKDAINTNGFTASTLDISNENIKARNFTFEQTKDAIAIVNHGLITVGKD 208
QY 350 GSQHLL---LDNKG-----HLFDIKSTATSYSLVHNSHPGEIKGLAQ 389
DB 209 GSNVLLGGKVKNEGVSIVNGSSILLAGQKITISDIINPTITYSIAAPENEAVALNGLDIPA 268
QY 390 AGTGSVSDG---KSKGISLGSCT---QSHNKTMLSQPEAHRSLTGTGWHQHPAGAARPQ 443
DB 269 KG-GNINVRATIRNOQKLSADSYSKDGKNIVLSAKEGA-----308
QY 444 GESIRLHDDKIHLHPGLVQWQSAKDTHSOLSRQADGKLYALKDNRTLIQ-----NLSDN 498
DB 309 -----EIGGVISA-----QNOQAKGKMLMTGDKVTLKTGAVIDLSGK 346
QY 499 KSEKLVDTKTSYVDQGVQVAILTDPGRHKMSIMPSLDASPEHSISLSLHFADAHQGL 558
DB 347 EGGETVL-----GGDERGE-----GKNGIOLAKKTSLEKGSTINVS-----382
QY 559 LHKSELEAQSVASHGRVLYVADSEGLFSAAIPKQGDGNCNELKMKAMPQHALDEHFHGDH 618
```

```
DB 383 --GK---ERGGRAIVMGDIALIDGN-----INAQSGDIAKTGGFV-----ETSGHD- 424
QY 619 QISGPFDDHGHQOLNALVKNNFRQOHACPLGNDHOFHPCGNLT-DALVIDNQGLHHTNPE 677
DB 425 ----LEIKDNAIVDA-----KWLLEDNFVNSAETAGRSNTSE 459
QY 678 PHEILDMHGSLALQEGKLYHFDQLTKWGTGASDCKOLKKGLDGAAYLLKDGVEKRLN 737
DB 460 DDEYTGSGNSASTPKRNK-----EKTTLTNTLES-----ILKKGTFVNIT 500
QY 738 INQ-----STSSIKHGTENVFSL-----PHVRNKPPEPDALQGLNKDDKAQAMAVIGN 786
DB 501 ANQRIYVNSINLSNGSLTLWSEGRSGGVVEINNDITTCDDTRGANL--TIYSGGWVDVH 558
QY 787 KYLALTEKEDI-----RSFOIKPGTQQLERPAQTLREGISGELKDIHVDH-----832
DB 559 KNISLGAQGNINITAKODIAFEKGSNQVITGGTIT-----SGNQKGFNNVNSLNGTSG 614
QY 833 -----KQNYALTHEGEVHPREAWONGAESSWH---KLALPQSES-----872
DB 615 LQFTTKRTNKYAITNKE-----GTLNISGKYNISMVLPKNESGYDKFKGRTYW 663
QY 873 KLKSLDMSHEHKPIATFEDGSOHOLKAGGWHAYAAPERGPLAVGT-----SGSQ 921
DB 664 NUTSLNVSSEGEFNLITD-----SRGSDSAGILTQPYNLNGISFNKD 705
QY 922 TVFN-----RLMGVKGKVIPIGSLTVKLSAQGTGAGTKVSKFSKFSERIRAYAFNPT 976
DB 706 TTVENVERNARVNFEDIKAPIGINKYSSLYAFNGISVSGGSDV-----FTLLASSSN 759
QY 977 MSTPRPTKNAAY-----ATOHGWQREGKLPKLYEQGALIKOLDAHNVHR 1021
DB 760 VQTPGVVINSKYFNVSTGSSLSRFTSGSTKGTGFSTEKDL--TLNATGGNITLLQVGTGD 817
QY 1022 NAPONDLOSLETLDLGEHGAELNDMKRFRDELEQSAT-----RSVTVLG---OHQ--- 1070
DB 818 MICKGIVAKKNITE---GGINITFGSRKAVTEIGENVITNNANVTLLIGSDFDNHQKPL 873
QY 1071 -----GVLKSNGEI-----NSEFKPSGKALVQSFNVNRSGQ-----1102
DB 874 TIKKDVIIINSGLTAGGNIVNIAGNLTVESNANFK-----AITNFTFNVGGILFDNKGNS 927
QY 1103 -----DLKSLQQAHA-----TPPSAESKLQSLM 1127
DB 928 NISIAKGGARFKIDNSKNLSITTNSSSTYRTIISGNITNKNGDLNITNMGSTEQ---984
QY 1128 GHFVSAGVDMSHQKEIPL-----GRQRPNDKTAULTKSLILDTVTIGE 1172
DB 985 -----IGDVSQKEGNTLSSDKINITKQITIKAGVDGENSDSDATNNANLTIKTKEKL 1039
QY 1173 LHEL-----ADKAKLVSDHKPDADQIKLRQOFTLREKYESNPVKHYTDMGFTHNKALE 1228
DB 1040 TQDLNISGFNKAETIA--KDGSDLTIGNTNSADGTNAKVTFOVKDKSKISADGKVKTLH 1097
QY 1229 ANYDAVKAFINAFKEHGVNLTTRTVLESQGSAAELAKKLKNTLL--SLDSGESMSFSRS 1286
DB 1098 S-----KVETSGSNNTED--SSDNNAGLTIDAKNVTVNNNITSHRAVSIAT 1143
QY 1287 YGGGVSTVFVPTLSK---KVPVPVIPA--GITLDRAYNLSPFSRTSGGLNVSPGRDGVGS 1341
DB 1144 -SGEITTKGTGINATTGNVEITAQTSILGGIESSGSVTLTATEGALAVS-----NIS 1197
QY 1342 GNIWVATGHVMPYMTGKTSAGNASDWLSAKHKISPDLRIGAASG-----TLOGTL 1394
DB 1198 GNTVTVTANSALATTLAGSTIKGTESVTTSSQ---SGD--IGGTISGGTVEVATESLTT 1252
QY 1395 QNSLKFKLTEDB-----LPGFIHGLTHGTLTPAELLQKGIHOMKOGSKLTFYSVDTSANL 1449
DB 1253 QNSKIKATTGEANVTSATGTIGGTISGN-----TVNVTANA 1289
QY 1450 -DLRAGINLNDGKPNGVTVARVSAG---LSASANLAAGSRERSTTISGQFGSTTSASNNR 1505
DB 1290 GDLTVG-NGAEINATPEGAATLTSSGKLTTEASSHTTS-AKGQVNLQAQDGSVAGSINAA 1347
```

QY 1506 PTEFNGVGA-----GANLTAALGV-----AHSSTHEGKPVGIFPAFTSTNVSAALALDNR 1556  
 Db 1348 NVTLNTGTLTTVKGSNNINATSGTLVINAQDAELNGAALGNHTVYNATNANGSGSVIATT 1407  
 QY 1557 SOSISLEKRAEPTVNDISELSTLGLKHKFKDSATTKML-----AALKELD 1602  
 Db 1408 SSRVNI---TGDLITINGLNIIS-----KNGINTVLLKGVKIDVKYIQPGIASVDEVI 1457  
 QY 1603 DAKPAQLHLOOHFSAKDVGVGDEREYAVRNKLKLVIR 1640  
 Db 1458 EAK-----RILEK---VKD-LSDEEREALAKLGSVAVR 1486

RESULT 11  
 US-08-728-470-2  
 ; Sequence 2, Application US/08728470  
 ; Patent No. 5928651  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Barenkamp, Stephen J  
 ; TITLE OF INVENTION: High Molecular Weight Surface Proteins  
 ; TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus  
 ; NUMBER OF SEQUENCES: 10  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Shoemaker and Mattare, Ltd.  
 ; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
 ; STREET: Bldg. 1  
 ; CITY: Arlington  
 ; STATE: Virginia  
 ; COUNTRY: U.S.A.  
 ; ZIP: 22202-0286  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/728.470  
 ; FILING DATE:  
 ; CLASSIFICATION: 424  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/302,832  
 ; FILING DATE: 16-MAR-1993  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US PCT/US93/02166  
 ; FILING DATE: 16-MAR-1993  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: GB 9205704.1  
 ; FILING DATE: 16-MAR-1992  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Berkstreser, Jerry W  
 ; REGISTRATION NUMBER: 22,651  
 ; REFERENCE/DOCKET NUMBER: 1038-633  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (703) 415-0810  
 ; TELEFAX: (703) 415-0813  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1536 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE:  
 ; US-08-728-470-2

Query Match 2.0% Score 190.5; DB 2; Length 1536;  
 Best Local Similarity 18.6% Pred. No. 0.00013;  
 Matches 297; Conservative 221; Mismatches 559; Indels 521; Gaps 72;  
 QY 304 VTPLAVTLDK-----GKIQALPAPNPALNTLLKQTLGKDTQHYLAH---HASSD 349  
 Db 149 INPNIITIGDALINTNGFTASTLIDISNENIKARNFTFEQTKDALAEIVNHGLITVGKD 208

QY 350 GSOHL---LDNKG-----HLFDIKSTATSYSVLHNSHPGEIKGLAQ 389  
 Db 209 GSVNLIGKVKNEGVISVNGSGISLLAGOKITISDIINPTITYSIAAPEAVNLGDI 268  
 QY 390 ACTGSVSDG---KSGKISLGSgt--OSHNTKMLSQPGEAHSRLTLTGIMQHPAGARPQ 443  
 Db 269 KG-GINVRAATIRNOGKLSADSVDKSGNIVLSAKEGA----- 308  
 QY 444 GESIRLHDDKIHLHPGLWOSADKTHSOLSRQADGKLYALKDNRTLQ-----NLSDN 498  
 Db 309 -----EIGGVISA-----QNOAGKGLMITGDKVTLTKTGAVIDLSGK 346  
 QY 499 KSSEKLVKIKSYSDQDQVAAILTDPGRHKMSIMPSLDASPEHSISLSLHFEADAHOGL 558  
 Db 347 EGGETYL-----GGDERG-----GKNGIQLAKKTSLEKGSINVS- 382  
 QY 559 LHGKSELEAQSVAISHGRVLVADSEKLFESAAPKOGDGNELKMKAMPQHALDEHFGHDH 618  
 Db 383 --GK---EKGGRVAVMGDIALIDGN-----INAGSGDIAKTGGFV-----ETSGHD- 424  
 QY 619 QISGFFHDDHGQNLALVKNFRQOHACPLGNDHQHPGWNLT-DALVIDNOLGLHHTNPE 677  
 Db 425 ---LFIKDNAIVDA-----KWLDFDNVYSINAEAGRSNTSE 459  
 QY 678 PHEILDMCHLGLSALQEGKLLHYFDQLTGWTGAESCKQLKGLDGAAYLLKDGVEKRLN 737  
 Db 460 DDEYTGSGNSASTPKRKN-----EXTLTNTTLES-----ILKGTFTVNIT 500  
 QY 738 INQ-----STSSIKHGTENVFSL-----PHYRNKPEPCDALOGLNKKDKKAQAMVIGVN 786  
 Db 501 ANQRIYVNSSINLSNGSLTLWSEGRSGGVEINNDITGDDTRGANL--TIYSGGVWDVH 558  
 QY 787 KYALTEKGTI-----RSFOIKPGTQQLERPAQTLSREGISGELKDIHVDH----- 832  
 Db 559 KNISLGAOGNINITAKQDIAFEKSGNOVITGOGTIT---SGNOKGFRFNNSVSLNLTGSG 614  
 QY 833 -----KONLYALTHEGEVFHQPREAWONGAESSSWH---KLALPOSES----- 872  
 Db 615 LOFTTKRTNKYAITNKF-----GTLNISGKVNISMVLPKNSGYDKFKGRTYW 663  
 QY 873 KLSLDMSHHEKPIATFEDGSOHQLKAGWHAYAAPERGLAVGT-----SGSQ 921  
 Db 664 NLTSLVNSESCEFNLTID-----SRGSDSAGTLTQPYNLINGISFNKD 705  
 QY 922 TVFN-----RLMOGVKGVIEGSLTVKLSAQTGOMTGAEGKRVSKSFSEIRAYAFNPT 976  
 Db 706 TTFNVERNARVNFIDIKAPIGINKYSSLYAFNGNISVSGGSDV-----FTLLASSSN 759  
 QY 977 MSTPRPIKNAAY-----ATOHWCQREGKLPLEYMOGALIKOLDAHNVHR 1021  
 Db 760 VQTPGVVINSKYFNVSTGSSLRFKYTSKTKGFSTIEKOL--TLNATGNTITLLQVEGTDG 817  
 QY 1022 NAPQPDLOSLETLDLGEHGAELNDMKRFRDELEQSAT-----RSVTVLG-----OHQ 1070  
 Db 818 MIGKIVAKKNITPE---GKNITFGSRKAVTEIEGNVTINNANVTLLIGSDFDNHQRPL 873  
 QY 1071 -----GVLKSNGEI-----NSEFKPSCKALVQSFNVNRSQ----- 1102  
 Db 874 TIKKDVIIINSGLTAGNIVNIAGNLTVESNANFK-----AITNFTNVGSLFDNKGNS 927  
 QY 1103 -----DLKSLQQAQVHA-----TPPSAESKLSQML 1127  
 Db 928 NISIAKGGARFKDIDNSKNLSITNSSSTYRTISGNTNKGDLNITNEGSDTEMQ--- 984  
 QY 1128 GHFVSAGVDMSHQKEIPL-----GRORDPNDKTALTAKSRILDTVTIGE 1172  
 Db 985 -----IGGDVSVQKEGNTLITSSDKINIKQITIKAGVDGENSDSDATNANLTIKTKEKL 1039  
 QY 1173 LHLEL-----ADKALVSDHKDPDQOIKOLROQFDTLREKRYESNPVKHYTDNGFTHINKALE 1228  
 Db 1040 TODLNSIGFNKAEITA--KDGSDLTIGTNSADGTNAKVTNFQVSKOSKISADGHKVTLH 1097

Qy	1229	ANYDAVKAFINAFKKEHHGVNLTTRTVVLBSOGSAELAKKKNVLL---SLDSESMFSRS	1286
Db	1098	S-----KVETSGSNNTED---SSDNAGLITDAKNNVTNNNTTSHKVASISAT	1143
Qy	1287	YGGGVSTVFVPTLSK--KVPVPVPGA-GITLDRAVNLFSRSTSGGVLNVPGRGGVS	1341
Db	1144	-SGEITTKGTGINATTGNVEITAGTGSILGGIESSGSVTLTATGALAYS-----NIS	1197
Qy	1342	GNIMVATGHDVMPYMTGKTSAGNASDMLSAKHKISPDLRIGAAVSG-----TLOGTL	1394
Db	1198	GNTVTVANSALTLTLAGSTIKGTESVTTSSQ---SGD-IGTISGGTVEVKATESLTT	1252
Qy	1395	ONSUKFKLTEDE----LPGFIHGLTHGTLTPAELLQKIEHQMKOGSKLTFPSVDTSANL	1449
Db	1253	QNSKIKATTGEANVTSATGTIGGTISGN-----TVNVTANA	1289
Qy	1450	-DLRAGINLNDGSKPNGVTARVSAG---LSASANLAAGSRERSTTSGFGSTTSASNNR	1505
Db	1290	GDLTVG-NGAEINATEGAATLTTSGKLTTEASSHTS-AKGVNLSAQDGSVAGSINAA	1347
Qy	1506	PTFLNGVA-----GANLTAALGV---AHSSTHECKPVGIPPAFTSTNVVSAALADNRT	1556
Db	1348	NVTLNNTGTLTVKGSINATSGTLVINAKDAELNGAALGNHTVVVNATNANGSGSVIATT	1407
Qy	1557	SQSISLELKRAPVPTSNDISELSTLTKKHFKDSATTKML-----AALKELD	1602
Db	1408	SSRVNI---TGDLITINGLNIIS-----KNGINTVLLKGVKIDVKYIQPGTASVDEVI	1457
Qy	1603	DAKPAEQLHILOQHFSAKDVGDEREYEAVRNLKKLVIR	1640
Db	1458	EAK-----RILEK-----VKD-LSDEEREALAKLGVSAVR	1486

RESULT 12  
US-08-617-697-2  
: Sequence 2, Application US/08617697  
: Patent NO. 5977336  
: GENERAL INFORMATION:  
: APPLICANT: Barenkamp, Stephen J  
: TITLE OF INVENTION: High Molecular Weight Surface Proteins  
: TITLE OF INVENTION: Of No. 5977336-typeable Haemophilus  
: NUMBER OF SEQUENCES: 11  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Shoemaker and Mattare, Ltd.  
: STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
: STREET: Bldg. 1  
: CITY: Arlington  
: STATE: Virginia  
: COUNTRY: U.S.A.  
: ZIP: 22202-0286  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: OPERATING SYSTEM: IBM PC Compatible  
: SOFTWARE: PatentIn Release #1.0, Version #1.30  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/617,697  
: FILING DATE: 01-APR-1996  
: CLASSIFICATION: 424  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 08/302,832  
: FILING DATE: 05-OCT-1994  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US PCT/US93/02166  
: FILING DATE: 16-MAR-1993  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Berkstresser, Jerry W  
: REGISTRATION NUMBER: 22,651  
: REFERENCE/DOCKET NUMBER: 1038-557  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (703) 415-0810  
: TELEFAX: (703) 415-0813  
: INFORMATION FOR SEQ ID NO: 2:

```

;
; SEQUENCE CHARACTERISTICS:
;
; LENGTH: 1536 amino acids
;
; TYPE: amino acid
;
; STRANDEDNESS: single
;
; TOPOLOGY: linear
;
; MOLECULE TYPE: DNA (genomic)
;
US-08-617-697-2

```

Query Match	2.0%;	Score	190.5;	DB 2;	Length	1536;			
Best Local Similarity	18.6%;	Pred. No.	0.00013;						
Matches	297;	Conservative	221;	Mismatches	559;	Indels	521;	Gaps	72;

QY	304	VTPLAVTLDK-----GKLOLAPNPALNTLLKOTLGKQDQHYLAH---HASSD	344
Db	149	INPNITIGKDAIINTNGFTASTLDSINENIKARNTFFOTKDALAEIVNHHGLUITVKGD	208
QY	350	GSOHLL---LDNKG-----HLFDIKSTATSYVLHNSHPGETKGRKLAQ	389
Db	209	GSVNLIGKVKNEGIVSYNGSGISLLAGOKIITISDIINPTITYSIAAPEAVNLGDIFA	268
QY	390	AGTGSVSDG---KSGKISLGSGT---QSHNKTMLSPQGAHRSLLTGCIWHPAGARPQ	443
Db	269	KG-GNINVRAATIRNOGKLSADSVSKDSCGNIVLSAKEGA-----	308
QY	444	GESIRLHDDKIHLHPGLVQWSADKTHSOLSRQADGKLYALKDNRTLQ-----	498
Db	309	-----EIGCVISA-----QNOAQGGKLMITCDKYTLKTGAVIDLSGK	346
QY	499	KSEKLVDKIKSYSDVRQGVAILTDPGRHKWMSIMPSLDASPESHLSLSLHFADAHQGL	558
Db	347	EGGETYL-----GGDERGE-----GKNQIOLAKKTSLEKGSTINVS-----	382
QY	559	LHGKSELEAQSVAISGRLVWADSECKLSAAIPKQGDGNELKMKAMPQHALDHFHGDH	618
Db	383	---GK---EKGGRAIVMGDIALIDGN-----INAQSGGDIAKTGGFV-----	424
QY	619	QISGFFHDDHGOLNALVKNFNFRQOHACPLGNHDHQFHPGWNLT-DALVIDNOLGHLHNTPE	677
Db	425	---LFIKDNAIVDA-----KEWLLDFDNVSINAETAGRSNTSE	459
QY	678	PHEIILDMGHLGSLALQEGKLHYFDOLTKGWTGAESDCQKLGKLDGAAYLLKDGVEVKRLN	737
Db	460	DDEYTGSGNSASTPKRKN-----EKTLTNTLTLES-----ILKGTFFVNIT	500
QY	738	INQ-----STSSIKHGENTVFSL-----PHVRNKPEPGDALOGLNKDDKAQAAVIGVN	786
Db	501	ANQRIYVNSSINLSNGLSLTWSEGRSGGGVEINNDITTDGTRGANL---TIYSGGWVDVH	558
QY	787	KYLATEKGD I-----RSFOIKPGTOOLRPAQTLRREGISGELKDIHVDH-----	832
Db	559	KNISLGAOGNINITAKQDIAFEKSGNOVITGQGTIT---SGNQKGFNFNVSLNGTSGG	614
QY	833	-----KQNIYALTHGEVFPHPREAWONGAESWSH---KLALPQSES-----	872
Db	615	LQFTYKTKNYAITNKE-----GTLNISGKVNISMWLPKNESGYDKPKGRTYW	663
QY	873	KLKSLDMSHEHKPIATFDGSOHLKAGGWHYAAAPERGPLAVGT-----SGSQ	921
Db	664	NLTSLVNSESGEFNLATID-----SRGDSAGTTLQPYNLINGISFNKND	705
QY	922	TYFN-----RLMOGVGKVIPIPGSLTVKLSAOTGMWGAECRKYSKFSERIRAFNPNT	976
Db	706	TTFNVERNARKVNFDIRAPIGINKYSSINLYASFNGNISVSGGGSVD-----FTLLASSSN	759
QY	977	MSTPPIKNAAY-----ATOHGWQREGLKPLYEMOGALIKOLDAHNVHRH	1021
Db	760	VQTPGVVINSKYFNVSNGSSILRPFKTSKGTGFSIEKDL---TLNATGNTITLLOVEGTDG	817
QY	1022	NAPQPDLOSKELEHLDLJDEHGAELLNDMKRRFDELEQSAT-----RSVTVLG-----	107
Db	818	MIGKGVIAKKNIITFE-----GGNITFGSRKAVTEIEGNNVITINNANVTLLIGSDFDNHOKPL	873

```

QY 1071 -----GVLKNGEI-----NSEKPSGKALVQSFVNRSG----- 1102
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 874 TIKKDWIISGNLTAGNIWNIAGNLVSNANFK-----AITNFTFNVGGLFDNKGNS 927
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1103 -----DLSKSLQAAVHA-----TPPSAESKLOSML 1127
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 928 NISIAKGGARFKDIDNSKLSITNSSTYRTIISGNITKNGDLNITNEGSDTFMQ--- 984
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1128 GHFVSAGVDSHOKGEIPL-----GRORDNDKALTAKSRILDTVTIGE 1172
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 985 -----IGGDSVQKEGNLTSSDKINIKQITIKAGVDGNSDSATNANITIKTELK 1039
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1173 LHEL-----ADKALVSDHKPDQIOLRQOFDTLREKRYESNPVKHYTDMGFTHKALE 1228
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1040 TODLNISGENKAEITA--KDGSDLTIGTNSADGTNAKKVTFNVOYKSKISADCHKVTLH 1097
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1229 ANYDAVKATINAPKKEHGHVNLTRRVLESQGSALAKKILNTLL--SLDSGESMSFSRS 1286
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1098 S-----KVETSGSNNNTED--SSDNNAGLTIDAKNVTVNNITSHKAVSISAT 1143
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1287 YGGGVSTVFPTLSK---KVPVPVPGA--GITLDRAYNLSFRTSGGLNVSFORDGVS 1341
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1144 -SCEITTKTGINATGNVEITAOGSIUGGIESGVSITLTATEGALAVS-----NIS 1197
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1342 GNMVATGHDVMPYMTGKTSAGNASDLSAKHKISPDLRIGAAVSG-----TLOGTL 1394
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1198 GNTVITVANSAGALTTLTLAGSTIKGTESVTTSSQ---SGD--IGGTISGTVKATESLTT 1252
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1395 QNSLKFKLTEDE-----LPGFIHLGHTLTPAELLQKGEHOMQKSKITFSVDTSANL 1449
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1253 QNSKIKATGGEANVTSATGTIGGTISGN-----TVNVTANA 1289
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1450 -DLRAGININEDGSKPNVGTARVAG---LSASANLAAGSRERSTTSQGFSTTSASNNR 1505
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1290 GDLTVG-NGAEINATGAATLTSSCKLTTEASSHTS--AKQVNLQAQDGSVAGSINAA 1347
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1506 PTFELNGVA-----GANLTAALGV-----AHSSTHEGKPVGIFPAFTSTNVSAALALDNR 1556
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1348 NYTLNTGTLTVKGSINATSTVLINAKDAELNGAALGNHTVWVATNANGSGSVIATT 1407
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1557 SOSISLELKRBPVTSNDTSELSTLTGKHKFSDATKML-----AALKELD 1602
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1408 SRVNI---TGDLITINGLNIIS-----KNGINTVLKGVKIDVYIOPGIASVDEVI 1457
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1603 DAKPAEQHLIOOHFSKADVDGDERVEAVRNLLKLVIR 1640
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1458 EAK-----RILEK---VKD-LSDEEREALAKLGVSAVR 1486
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

## RESULT 13

US-08-617-697-10

Sequence 10, Application US/08617697

Patent No. 5977336

GENERAL INFORMATION:

APPLICANT: Barenkamp, Stephen J

TITLE OF INVENTION: High Molecular Weight Surface Proteins

TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus

NUMBER OF SEQUENCES: 11

CORRESPONDENCE ADDRESS:

ADDRESSEE: Shoemaker and Mattare, Ltd

STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202-0286

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/617,697

```

; FILING DATE: 01-APR-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 05-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-557
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1600 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-617-697-10

```

```

Query Match      2.0%; Score 187; DB 2; Length 1600;
Best Local Similarity 19.1%; Pred. No. 0.00024;
Matches 328; Conservative 234; Mismatches 632; Indels 520; Gaps 82;

```

```

QY 304 VTPLAVTLTK-----GKQLAPDNPPALNTLLKTLCKDTHOYLAH---HASSD 349
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 149 INPGITIGKDAIINTNGFTASTLDISNENIKARNETLEOTKDKALAEIVNHGLITVGKD 208
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 350 GSOHLL---LONKG-----HLFDIKSTATSYSVLHNSHP-----GEIKG 385
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 209 GSVNLIGCKVKEGIVSGVSGISLLAGOKITISDINPTIYSIAAPENAINLGDIFA 268
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 386 KLAOAGTGSVSDGKSGKISLGSMT--QSHNKTMLSQPGFAHRSLLTGIWQHAGARPQ 443
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 269 KGGNINVRATIRNK--GKLSADSVSKDSGNIVLSAKEGA----- 308
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 444 GESIRLHDDKIHILHPELGVWOSADKTHSLSQSRADGKLYALKDNRITLQ-----NLSDN 498
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 309 -----EIGGVISA-----QNOQAAGGKLMITGDKVTLTKTGAVIDLSGK 346
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 499 KSSEKLVDKIKSYSDORQOAVAILTDPORHKMSIMPSLDASPESHISLSLHFADAHQGL 558
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 347 EGGETYL-----GGDERGE-----GKNGIQIAKKTTLKKGSTINVS----- 382
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 559 LHGKSELEAQSVASHGRVLVADSEGKLFSAAIKPGDGNELKMKAMPQHALDEHFGHHDH 618
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 383 --GK--EKGGRAIVWGDIALLIDGNINAGSIDIATG-----GFVETSGHDL 424
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 619 QISG-----PFHDDHGOLNALV--KNNFROQHACPLGNDHOFHPGWNLTDALVIDN 667
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 425 SIGDDVIVDAKEWLLDPDDVSIETLTSGRNNTGENOYTTGDTKESPKGNSISKPTL-- 482
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 668 QLGLHHNPEPEHILDMCHL-----GSLAQEGKLYHFDLTKGTGCAESDCK 715
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 483 -----TNSTLEQILRRGSYVNITANNRIYVNSNLSNLSL----- 518
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 716 QLKKGLDGAAYLLKDGVEVKRLNINOSTSIKHTGENTVFSPLHV---RN-----KPEP 764
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 519 -----TLHTKRDG-----VKINGDITSNENGLTIKAGSVWDVHKNTLTGTGFLNIVA 566
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 765 GDALQGLNKDDKAQ-----AMAVIGVNYKYLALTEKDIRSFQIKPGTQOLERPAOTL 816
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 567 GDSVAFEREGDKARNATDAQITAQGTITVVK-----DDKQFR-----F 604
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 817 SREGISGELKDIHYDVKONLYALTHEGEV-----FHQPRE---AWONGAESSSHKKLA 866
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 605 NNVSNGTGKGLKFIANNONNETHKFDINISGIVTINOTTKKQVYWNASKDSYWNVSS 664
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 867 LPOSESKLSLDMSHKPIATFTFDGSOHLKAGWHAYAAPERPGLAVTSGSTVFNRR 926
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```



```
Db 665 L-----TLNVQKFTPIKFDVDSNGQDLRSSRRSFAGVHNGI-----CGKTFN- 710
QY 927 LMQGVKGVI-----PGSGTVKLSAQGTGTMGABGRKVSXKFSRIRAYAFNPT 976
Db 711 --IGANAKALFKLKNAAADPKKELPITFNITA-TGNSDSSVMFDIHANLTSRAAGIN 767
QY 977 MSTPRPKNAAYATOHGWOGREGLKPLYEMOGALIKOLDHNVHNA--POPDLQSKLET 1034
Db 768 MDS-----INITGGLDFEITSITSHNRNSNAFEIKDL-----T 798
QY 1035 LDLGHEGAELNDKMRFRDELOSATRS---VTVLGHQGVLSKNGSEINSEFKPSGKAL 1091
Db 799 INATGSNFSKOTKOSFYNEYSKHAINSSHNLTLG---GNVLGGE-NSSSSITGINI 854
QY 1092 VOSFNV-----NRSQDLSK-----SLQQAHVHATPPSAE 1120
Db 855 TNKANVTLQADTSNSNTGLKKRRTLTLGNTSVEGNLSLTGANANIVGNLSIAEDSTFKGEA 914
QY 1121 SKLQSLMGLHFVSAGVDMSHQGEIPLGRQDPNDKTAL-----TKSRLILDVT--I 1170
Db 915 SONLITGFTTNGTANINIKGVVKLG---DIANKGGLNITTNASGKTIIINGNITNEK 971
QY 1171 GELHELADRAKLVDHDKPDADQIKLRQOPDTLREKRYESNPVKHYTDMGFTHNKALEAN 1230
Db 972 GDLN-----IKNIKADA-----EIQIGGNISQKEGNLTSSDKVNITN 1009
QY 1231 YDAVAFINAFK---KEHGVNLTTRT-VLESQGSABELAKKLNTLLSLDSGESMSFSRS 1286
Db 1010 QITIKAGVEGRSDSEAEANANITQTKELKLAGDLNIS-GFNKAEITAKNGSOLDITGNA 1068
QY 1287 YGGGVSTFVPTLSKKVPVPIGAGITLDRAYNLFS---RTSGGLNVSFGRDGVSGN 1343
Db 1069 SGNAD-----AKVTPDKVDSKISTD-GHNVTLNSEVKTSG-SNAGNDNS-TGL 1118
QY 1344 INVATGHDVPMYMGKKT-----SAGNASDWLSAKHISPDLRTGAAGVSGTQGLQON-S 1397
Db 1119 TISAKDVTNNVTSHKTINISAAAGN---VTTKEGTTIN-----ATTGSEVETAQNGT 1169
QY 1398 LAFKLTDELPGFIHGLTHGLTPAEILLQKIEHOMQGSKLTFSDVTSANLDRAGI-- 1455
Db 1170 IKGNTISQNV-----TVTATENLVT-TENAVINATSGTNISTKTG-DIKGGIES 1217
QY 1456 ---NLNEDGSKPVGTVARVSAGLSASANLAAGSRERSTTSQGFSTTSASNNRPTFLNGV 1512
Db 1218 TSGNVNITAS---GNTLKVSNTQDQVTVTADAGALTFTA---GSTISATT----- 1262
QY 1513 GAGANLTAALGVASHSTHEKPGVIFPAFTSTNVSAALALNRTSQSISLELKRAEPTVS 1572
Db 1263 -GNANITTTKTGDIN-----GKVESSSGSVTLVATGATLAVGNISGNTVTI----- 1307
QY 1573 NDISELSTLTKGHFK--DSATTMKLAALKELDDAKPABQLHILQOHFSKADV-----VGD 1625
Db 1308 ADSCKLTSTVGSTINGTNSVT-----SSQSGDIEGTISGNTVNVNTASTGD 1353
QY 1626 ERYEAVRNKLVIRQOAAADSHSMELGSAHSTYNNLSRINNDGIVELLKHFDALPA 1685
Db 1354 ---LTIGNSAKVEAKNGAA-TLTAESGKL---TTQTGSSITSSNGQTTLTAK--DSSI-A 1403
QY 1686 SSAKRLGEMMNNDPALKDI-IKOLQSTPFSSASVSMELKDLGRLQTEKAILDGKVGREEV 1744
Db 1404 GNINAAVNLTNTGTLTTGDSKINAT---SGTLFINAKD-----AKLDGAAS---- 1448
QY 1745 GVLFQDRNNLRVKSVSQSVSKSEGFNTPALLLGTSNSAAMSMERN-IGTFNKYQDQ 1803
Db 1449 ---GDRTVVNATNASGSGNVTA-----TSSSVNITGDLNTINGLNIISENG 1493
QY 1804 NTPRRFTLEGIAQANFOVASALTDLKEGLEMK 1837
Db 1494 NTVLRGRKEIDVKYIQPCVASV-----BEVTEAK 1522
```

RESULT 14

```
US-08-328-254-6
; Sequence 6, Application US/08328254
; Patent No. 5710022
; GENERAL INFORMATION:
; APPLICANT: Zhu, Xueliang
; APPLICANT: Lee, Wen-Hwa
; TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/328.254
; FILING DATE: 24-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/141,239
; FILING DATE: 22-OCT-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-CJ 1191
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2482 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-328-254-6
```

```
Query Match 2.08; Score 186.5; DB 1; Length 2482;
Best Local Similarity 19.3%; Pred. No. 0.00053;
Matches 288; Conservative 245; Mismatches 592; Indels 369; Gaps 75;

QY 488 DNRTLQNLSDNKSSEKLVKIKSYSDQKGOVAILTDTGPRHKMSIMPSLDASPEHSL 547
Db 4 EKENLQSKINHELTCLKTQOIKSHEYNER-----VRLTLEMDRE---NL 43
QY 548 SLHFADAHQGLLHGKSELEASVALSHGRKLVVADSEGK-----LFSAAIPKQDGN 598
Db 44 SVEIRNLHNLVDSKSVEVETQKLAYMELQQAFAFSQKQKEJENMCLKTSQLTQGVDEL 103
QY 599 ELKMKAMPOHALDHFGHDHQSFGFFHDDHGNALNVLKNFRQOHACPLGNDHQFHPGWN 658
Db 104 EHKQLLSNEIMDKDRCTQD-----LHAYESLRDLKS---KDALSVTNEHQ----- 149
QY 659 LTDAVLIDNQLGLHHTNPPEHILDMGLGSLALQEGKLHY-PDQITKGTWGAESDCKQL 717
Db 150 -RSLAFDQQAPMHHSFAN-----IIGEGSMPSESRSECRLEADQSPKNSAILQNRVDSL 203
QY 718 KKGIDGAAYLLKD-----GEVKE-----LNINQSTSSIKHGT--- 749
Db 204 EFSLESQKMSDLQKQCFELVOIKGEIEENIMKAEQMHQSFAVETSQRIKQEDTSAH 263
QY 750 ENVE--SLPHVRNKPPEGDALQGLNKDKKAQAMAVIGVKNKYLAITEKDIRSFQIKPQTQ 807
Db 264 QNVVAETLSALENKEK---ELQLLNDKVETEQAIEQLKSNHLE-DLSELQLLSETL 319
QY 808 QLE-----RPAQTLREGISGELKLTVDHDKQNLVYALTHEGEVPH----- 847
```

Db 320 SLEKKEMSSIIISLKNKEIEELTOE--NGTLKEINASLNQEKWNLIQKSESFANYIDREK 377  
QY 848 -----OPREAWONGAESSSWHKLAIPQSESKLKLSDMSHEHKPIATFE 890  
Db 378 SISELSDOYKOEKLIILQRCETGNAVYEDLSQYKAAQEKNSKLECL--LNECTSLCENR 435  
QY 891 DGSOHOLKAGGWHAYAAAPER--CPLAVGTSGSQTVEFNL---MQGVKGKVIPIGSGGLTVK 944  
Db 436 KNELEQLK---EFAKEHQEFLTKLAFABERNQNLMLLETVOALRSEMTDNQ---N 488  
QY 945 LSAQTGGM-----TGAEGRKVSSEKSERIRA-YAFNPTWST-----PRPIKNAAYA 989  
Db 489 SKSEAGGLKOEIMTLKEQNKMKQKENDLLQENBQLMKVMKTHCEQCNLESEPIRNSV-K 547  
QY 990 TQHCWQREGKLPYEMOGALIKO--LDAHNVHRNAPQPDLOSKELETLDLGEHGAELL-N 1046  
Db 548 ERESERNOCNFKPOMLE---VKEISLDSYNAOLVQLEAMLRNKLQSEKECEKLOH 604  
QY 1047 DMKRFDELQESATRSVTYLGQHQGVLSKNGEINSEFKPSFG-KALVQSFNVNRSQDLS 1105  
Db 605 ELQITRGDLTSLNODMQ--SOEISGLK-DEIDAEBEKYISGPHELSTSQNDN---AHLQ 658  
QY 1106 KSLQ-----QAVHATPPSAESKLQMLGHFVAGVDMSHQGEIPIGRORDP 1152  
Db 659 CSLOTTMKNLNELEKICEILOAEKYELVTELDNDSRSECITATRKMAEVEGKL-LNEVKIL 717  
QY 1153 NDKTALTKSRLILDVTIGELHELADKAKIVS-----DHKPDADOIKOLRQOFDT 1202  
Db 718 NDDSGLLHCELYED-IPGEGFGEQNEQHPVSLAPIDESNSYEHILTSD--KEVQMHFAE 774  
QY 1203 LREKYENPVKHYYTDMGTHNK--ALEANYDAVKAFINAFKKEHGVNLTTRTVLES-Q 1259  
Db 775 LOEK-FLSLOSEH---KILHQHCOMSSKSELQTVYVDSLKAE---NLVLTNLRNFQ 825  
QY 1260 GSAELAKLKNLTLLSDGESMSFSRSYGGYSTVFPVPLSKVPVPIPGAGITIDRAY 1319  
Db 826 G--DLVKEMQ---LGLLEG-----LVPSLSSC-VPDSSSLSLGDSFF 863  
QY 1320 NLSFRTSGLVNFGRDGVSGN-----IMVATGHDVMPYMTGKTKSAGNASD----- 1368  
Db 864 YRALLEQTDGMSLLNLEGAVSANQCSVDVEFCSSIQE--ENLTRKETPSAPAKGVEELE 921  
QY 1369 ---WLSAKHKISPOLRIGAAVSGTLQTLONS-----LKFPLTE 1404  
Db 922 SLCEVYROSLEKEKME-----SOGIMKNKEIQEQLLSSERQELDLRKOYLSF 973  
QY 1405 DELPGFIHGLTHGTTPAEILLQGHIEHOMKQSKLTFSDTSANLDRAGINLN----- 1458  
Db 974 NE--QWQOKLTSVTL---MESKLAEEKQTEQLSLELV-ARLOLQ-GLDLSSRSLLG 1025  
QY 1459 ---EDGSPNGVTARVSAGLSASA-----NLAAGSRERSTTSQ 1494  
Db 1026 IDTEAIOGRNESCDSIKETSETTERTPKHDVHOICDKDAQODLNL---DIEKITETGA 1082  
QY 1495 FGSTTSASNNRPTFLNGVAGANLTAALGAHSSHTECKPGVIFPAFTSTNVSAAALDN 1554  
Db 1083 LKPTGECGSGSDPTNYEPPGEDTKOGSSECISEL-----SFGFNALVPMDFTLG 1132  
QY 1555 RTSOSISLELKRAPVTSNDISELTSTLQKHFKDSATTMKMLAALKELDDAKPAEOLHILQ 1614  
Db 1133 NQEDITHNLQLRVKE--TSNENLRLLHVI--EDDRKRVESLLNEMKELD-----SKLHLOE 1183  
QY 1615 QHFSAKQDVGDREYAVRNKLKV--IRQQAAD--SHSMELGSAHSHTTYNNLSRI--NNDG 1670  
Db 1184 VOLMTK-----TEACTEIKIVGELKENSIDLSEKLEYFSCDHQEL---LQVETSEG 1233  
QY 1671 IVELLKHFDAAALPASSAKRLGEMMN--ND--PALKDIILOQOSTPSSASVSME--L 1722  
Db 1234 LNSDLEHAD---KSSREDIGDNVAKVNDWKEKRFLOVENELSRIRSEKASIEHEALYL 1289  
QY 1723 KQGLR-EOTEKAILD-----GKV---GREEGVGLFQDRNNLRKVSVSQSVS 1766  
Db 1290 EADLEVQVOTEKLEKDNENKQKIVICVLEBELSVVTSERNQLRGELDTMSKTT 1343

RESULT 15  
US-08-353-700-1  
; Sequence 1, Application US/08353700  
; Patent No. 5599919  
; GENERAL INFORMATION:  
; APPLICANT: YEN, TIMOTHY J.  
; APPLICANT: RATTNER, JEROME B.  
; TITLE OF INVENTION: NUCLEIC ACID ENCODING A  
; TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,  
; TITLE OF INVENTION: AND METHODS OF USE  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: DANN, DOREMAN, HERRELL AND SKILLMAN  
; STREET: 1601 MARKET STREET, SUITE 720  
; CITY: PHILADELPHIA  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19103-2307  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/353,700  
; FILING DATE: 09-DEC-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: REED, JANET E.  
; REGISTRATION NUMBER: 36,252  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (215) 563-4100  
; TELEFAX: (215) 563-4044  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3248 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: HUMAN  
; US-08-353-700-1

Query Match 2.0%; Score 184.5; DB 1; Length 3248;  
Best Local Similarity 18.4%; Pred. No. 0.0011;  
Matches 363; Conservative 292; Mismatches 759; Indels 563; Gaps 89;  
QY 108 RDLIAR-----DDGETQHEAAAPDAARLTRSGVKKRRNDDMAGRPVMVKGSGGDKVPQTQ 162  
Db 340 RDELVRTTAQVDOASTKYTALEQKLKLTEDLSCQRQNAESAR-----CSLEQKIKKK 392  
QY 163 OKR-----HQLNFGQMRQTMLSKMAHPASANAGDRLOHSPHPHPPGSHHEIKKEEVPVGT 216  
Db 393 EKEFQELSRQORSFQTLDOECIQMKA-----RLTQELQQAQKNNHNVLOAELDKLT 443  
QY 217 SKATTAHADRVETIAQE-----DDDSEFOQLHOORLARENPPOPKLGVATPISARFQPK 272  
Db 444 SVKOOLENNLEEFKQKLCRAFOAQAQSOIKENELRSMEEKKENNL-----LKSHSEQK 498  
QY 273 LTAVAESVLGTDYDTQSPKPSQSMKLGSGAGVTP-----AVTLDKGLK 316  
Db 499 AREVCHLEAELKNTKQCLNQSNQFNAEEMKAKNTSQETMLRDLQEKINQOENSLTLEKLL 558  
QY 317 QLAPDNPALNTLLKQTLGNDQTHYLAHHASSDGSQHLILLDNKGHLFDIKSTATSYSLVH 376  
Db 559 AVA-----DLEKQ-----RDCSDLLKKRHHI----- 581

QY 377 NSHPGEIKGLAQAGTGSVSDCKSGKISLGSQTOSHNTKMTLSQPGFAHRSLLTGIWHP 436  
Db 582 ---EQLNDKL-----SKTESKALLS----- 600  
QY 437 AGAARPOGESIRLHDDKIHLPE---LGVWOSADDTSHSQRQADGKLYALKDNRTLO 493  
Db 601 ---ALEKKKEVEELKEEKTFLSCWKSENEKLLTQMESE-----KEN-LQ 641  
QY 494 NLSDNKSEKLVKIKSYSDVQGOVAILTDTGPRHKMSIMPDLADSPESHISLSLHAP 553  
Db 642 SKINHLETLCTKQIKSHEVNER-----VRTLEMDRE---NLSVEIRN 681  
QY 554 AHQGLLHCKSELEAQSVALSHGRVLVADSEK-----LPSAIPKOGDGNELKMA 604  
Db 682 LHNVLDSKSVEYETOKLAYMELQOKAEFSDQRHOKKEIENMCLTQSQTQGVDELEHLQL 741  
QY 605 MPOHALDEHFGHDHQTSGFFHDDHGLNALVKNFNQHQHACPLGNDFHOFPGWNLTDALV 664  
Db 742 LSNEIMDRDRCYD---LHAYESLRDLLKS---KASLVTNEDHQ-----RSLLA 786  
QY 665 IDNQLGLHTNPPEPHILDMHGLSLALQEGKLYH-FDQLTGWTGAESDCQKQKKGIDG 723  
Db 787 FDOQPAMHHSFAN-----TIGEGQSPSEKSECRLEADQSPKNSAILQNRVDSLEFLES 841  
QY 724 AAYLLKD-----GEVNR-----LNINQSTSSIKHGT---ENVP-- 753  
Db 842 QKOMNSDLQKQCEELVQIKGEIEENLMKABQMHSQFVAETSORISKQEDTSAHQNVAAE 901  
QY 754 SLPHVRNKPPEPDALQGLNKKDKAQAAMAVIGVNKYLATTEKGDITRSQIKPCTQOLE--- 810  
Db 902 TLSALENKEK---ELQLLNDKVETEOAEIOELAKKSNHLE-DSUKELQLLSETLSLEKKE 957  
QY 811 ---RPAOTLSREGISGELKDTIHDVHKONLYALTHTHEGEVTH----- 847  
Db 958 MSSIISLNKREIEELTQOE---NGTLKEINASLNQEKMLQKSESFANYIDREKSISELS 1015  
QY 848 ---QPREAWQNGAESSSWHKALPQSESKLKSLDMSHEHKPIATFEDGSHQ 896  
Db 1016 DOYKQEKLIILQCEETGNAYEDLSQYKAAQKNSKLECL--LNECTSLCENRKNLEQ 1073  
QY 897 LKAGGHAYAAPR---GPLAVGTSGSOTVFNRL---MOGVKGKVPISGSLTVKLSAOTG 950  
Db 1074 LK---EAFAKEHOFETKLAFAERNQNLMELETVOQALRSEMTDNQN---NSKSEAG 1126  
QY 951 GM-----TGAEGRKVSSKFSERIRA-YAFNPTMT-----PRPIKNAAYATQHGQW 995  
Db 1127 GLKQEIWTLKEEQNKQKEVNDLIQENELQMLKVMKTKHECQNLSEPIRNSV-KERESER 1185  
QY 996 GREGLAPLYEMOGALIKO---LDHNVHRNAPOPDLOSKLETIDLGEHGAELL-NDMKRFR 1052  
Db 1186 NQCNFRQMDLE---VKEISLDSYNAQLVQLEAMLRNKLQSEKKEKELQHELOTIR 1242  
QY 1053 DELEQSATRSVTVLQGHQGVLSKNGEINSEPKSPG-KALVQSFNVNRSGDLSKSLQ-- 1109  
Db 1243 GDLETSNLQDMQ--SQEISGLK-DEIDAEEKYISGPHELSTSQDN---AHLQCSLQTT 1296  
QY 1110 ---QAVHATPPSAESKLSQMLGHFVSAGVDSHQKGEIPIGRQRPNDKTKAL 1158  
Db 1297 MNKLNELEKICEILQAEKYELVTELNDRSRSECIATRKMAEVEGKL-LNEVKILNDDSG 1355  
QY 1159 TKSRLLIDTVTIGELHELADKAKLYS-----DHKPADQIKOLRQOFTDLREK-- 1206  
Db 1356 LHGELVED-IPGFGFQPNQEPHFVSLAPLDESNSYEHLLTSD---KEYOMHFAELQEKEL 1412  
QY 1207 ---RYESNPVKHYTDMGFTHNKALEANYDAVKAFTNAFKKEHH--- 1246  
Db 1413 SLOSEHKILHDQCMSSKSELOTYVDSLAKAENLVLTN---LRNFQGLVYKEMOLGLE 1469  
QY 1247 ---GVNLTTRTVLESQSAELAKKKNLT-----LSLD----- 1276  
Db 1470 EGLVPSLSSSCVPDSSLSLSDSGSFYRALLEQTGDMSLLSNLEGAVSANQCSVDEVFCS 1529  
QY 1277 ---SGESMSFS---RSYGG-----GVSTVFVPTLSKKVPVPVPIPGAGITILD 1316

Db 1530 SLQTYVDSLKAENLVLTNLRNFQGLVKEMQLGLEGLVPSLSSC-VPDSSSLSLD 1588  
QY 1317 RAYNLSFSSTSGGLNVSFGRDGVSGN-----IMVATGHVMPYMTGKTKTSAGNASD-- 1368  
Db 1589 SSFYRALLEQTGDMSLLSNLEGVVSANQCSVDEVFCSLQE--ENLTRKETPSAPAKGVE 1646  
QY 1369 ---WLSAKHKITSPDLRIGAAVSGTLQGTLONS-----LKP 1401  
Db 1647 ELESICEVTRQSLKELEKME-----SQIMKNKEIOLEQLLSERQOELDLRQY 1698  
QY 1402 LTEDELPFGTHGLTHTPAELLQKQIEHQMKQSGKLTFSVDTSANLDRAGINLN--- 1458  
Db 1699 LSENE--QWQOKLTSVTLE---MESKLAEEKKQTEQLSLELV-ARLQLQ-GLDLSRS 1750  
QY 1459 ---EDGSKPNGVTARVSAGLSASA-----NLAAGSRERSVT 1491  
Db 1751 LIGIDTEDATQGRNESCDISKEHSETTERTPKHDVHQCIDKDAQODNLN---DIEKITE 1807  
QY 1492 SGQFGSTTSASNNRPTFLNGVGAGANLTAALGVVAHSTHEGKPVGIFPAFTTNNVSAALA 1551  
Db 1808 TGAVKPTGECSGEQSPDTNVEPPGEDKTQSGSEICSEL-----SFGPNALVPM 1857  
QY 1552 LNRTSQISLELKRAEPTVTSNDISELTSTLGKHFKDSATTKMLAALKELDDAKPAEQ 1611  
Db 1858 FLGNQEDIHNLQURVKE--TSNENLRLLHVI--EDDRKVESLLNEMKELD-----SKLH 1908  
QY 1612 ILQOHTSAKDVGVDERYEAVRNKLV--IRQOAAAD-SHSMELGSAHSTTYNNLSRI-N 1667  
Db 1909 LOEVQLMTK-----TEACIELEKIVGELKKENDSEKLEVFSCDHQEL---LQVET 1958  
QY 1668 NDGIVELLHKKHFDALPASSAKRIGEMMN--ND---PALKDIKQLOSTFFSSASVSME- 1721  
Db 1959 SEGINSOLEMHAD---KSSREDIGDNVAKVNDMSKRFVDYENELSRIRSEKASIEHEA 2014  
QY 1722 --LKDGLR-EQTEKAILD-----GKV---GREEVGVLFQDRNNLRKVSVSQSVS 1766  
Db 2015 LYLEADLEVQTEKLCLEKDNENKOKVIVCLEELSVVTSERNOLRGELDTMSKKT 2071

Search completed: June 5, 2001, 18:16:59  
Job time: 185 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2001, 18:13:14 ; Search time 55.68 Seconds  
(without alignments)  
1886.962 Million cell updates/sec

Title: US-09-596-784-2

Perfect score: 9448

Sequence: 1 MELKSLGTEHKAHVHTAAHN.....NPQVASALTLKKEGLEMK 1838

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_0401.\*  
1: /SID56/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SID56/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SID56/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SID56/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SID56/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SID56/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SID56/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SID56/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SID56/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SID56/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SID56/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SID56/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SID56/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SID56/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SID56/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SID56/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SID56/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SID56/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SID56/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SID56/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SID56/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SID56/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9448	100.0	1838	20 W98011	Hypersensitive res
2	9448	100.0	1838	21 Y71095	Erwinia amylovora
3	9448	100.0	1838	21 Y84856	A hypersensitive r
4	1041	11.0	201	21 Y69258	Amino acid sequenc
5	275	2.9	2514	21 Y75097	Neisseria meningit
6	245	2.6	2599	21 Y75098	Neisseria meningit
7	241.5	2.6	1837	21 Y85564	Human homologue of
8	232	2.5	1981	19 W42634	Protein sequence t
9	225	2.4	2608	21 Y85574	Hs-UNC-53/3/GFP f
10	224.5	2.4	2541	21 B41087	Human ORF851
11	222	2.3	2385	21 Y85569	Human homologue of

12	221.5	2.3	2048	21 Y75096	Neisseria gonorrhoe
13	210	2.2	3647	11 R05041	Flamentous haemag
14	207.5	2.2	3596	21 Y87407	Bordetella pertuss
15	204.5	2.2	1992	17 W04505	Moraxella 200 kDa
16	203.5	2.2	1780	19 W53863	Human gravin polyp
17	203.5	2.2	1780	21 B15380	Human gravin prote
18	203.5	2.2	1978	20 Y27230	Amino acid sequenc
19	201	2.1	2353	17 R99393	Haemophilus adhesi
20	198	2.1	1968	12 R10941	Mutant protease (d
21	197.5	2.1	1962	21 B29694	Mouse FLASH protei
22	197	2.1	1962	12 R10558	Mutant protease (A
23	197	2.1	2272	18 W21731	GAL4/HA/NuMA fusio
24	195	2.1	2326	21 Y71158	Rat phosphodiester
25	194	2.1	2042	19 W56319	Haemophilus paraga
26	194	2.1	2954	20 Y01632	Amino acid sequenc
27	192	2.0	1974	12 R10940	Mutant protease (d
28	191	2.0	1962	12 R10561	Mutant protease (N
29	191	2.0	1962	12 R10557	Mutant protease (A
30	191	2.0	2192	18 W21732	LexA/NuMA fusion p
31	190.5	2.0	1536	14 R41723	High molecular wei
32	190	2.0	2411	21 B23860	Haemophilus influe
33	189.5	2.0	1536	14 R41725	High molecular wei
34	189.5	2.0	1536	18 W30293	Non-typeable Haemo
35	189.5	2.0	2115	21 Y49937	Human NuMA protein
36	189	2.0	1931	21 B23258	Human apoptosis in
37	188.5	2.0	1420	20 W81025	AIb1 (Amplified in
38	188.5	2.0	1522	20 Y21975	Human steroid rece
39	188.5	2.0	1536	15 R63505	Haemophilus high m
40	188.5	2.0	1536	21 B01846	Haemophilus influe
41	188.5	2.0	1962	12 R10560	Mutant protease (K
42	188.5	2.0	1962	12 R10563	Mutant protease (K
43	188.5	2.0	2039	19 W56322	Haemophilus paraga
44	188.5	2.0	2442	21 Y77575	Human cytoskeletal
45	188.5	2.0	2517	21 Y71159	Human phosphodiect

#### ALIGNMENTS

RESULT 1  
W98011  
ID W98011 standard; Protein; 1838 AA.  
XX  
AC W98011;  
XX  
DT 21-JUN-1999 (first entry)  
XX  
DE Hypersensitive response elicitor DspE.  
XX  
KW Hypersensitive response elicitor; DspE; disease resistance;  
KW insect resistance; biological control; transgenic plant.  
XX  
OS Erwinia amylovora.  
XX  
PN WO9907206-A1.  
XX  
PD 18-FEB-1999.  
XX  
PF 24-JUL-1998; 98WO-US15426.  
XX  
PR 06-AUG-1997; 97US-0055105.  
XX  
PA (CORR ) CORNELL RES FOUND INC.  
XX  
PI Beer SV, Bogdanove AJ, Kim JF, Wei Z;  
DR WPI; 1999-180362/15.  
DR N-PSDB; X24810.  
XX  
PT Nucleic acid encoding hypersensitive response-eliciting protein -  
PT used to improve growth of plants and impart resistance to disease  
and insects  
XX

PS Claim 18; Page 51-56; 75pp; English.

XX This polypeptide comprises the 198 kDa hypersensitive response

CC elicitor protein DsPE of *Erwinia amylovora*. The nucleotide

CC sequence of the dsp region of *E. amylovora* strain Ea321 was

CC determined using subclones of pCP430. A two-gene operon was

CC discovered comprising dspE (see X24810) and dspF (see X24811). The

CC isolated dsp DNA molecules and encoded proteins can be used to

CC impart disease resistance to plants, to enhance plant growth,

CC and/or to control insects on plants. This is achieved by applying

CC a hypersensitive response elicitor protein or polypeptide in a

CC non-infectious form to plants or plant seeds, or by producing

CC transgenic plants or plant seeds transformed with DNA encoding a

CC hypersensitive response elicitor. Protection can be provided

CC against a wide range of viruses, bacteria, fungi and insects, e.g.

CC tobacco mosaic virus and tomato mosaic virus, *Pseudomonas syringae*,

CC *Xanthomonas campestris*, *Fusarium oxysporum*, *Phytophthora infestans*,

CC armyworm, diamondback moth, etc. The method avoids use of

CC infectious agents or polluting chemicals. Claimed transgenic

CC plants are selected from alfalfa, rice, wheat, barley, rye, cotton,

CC sunflower, peanut, corn, potato, bean, pea, chicory, lettuce,

CC endive, cabbage, brussel sprout, sweet potato, beet, parsnip,

CC turnip, cauliflower, broccoli, turnip, radish, spinach, onion,

CC garlic, eggplant, pepper, celery, carrot, squash, pumpkin,

CC zucchini, cucumber, apple, pear, melon, citrus, strawberry, grape,

CC raspberry, pineapple, soybean, tobacco, tomato, sorghum, sugarcane,

CC Arabidopsis thaliana, Sainpaulia, petunia, pelargonium, poinsettia,

CC chrysanthemum, carnation and zinnia.

XX

SQ Sequence 1838 AA;

Query Match 100.0%; Score 9448; DB 20; Length 1838;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1838; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MELSLGTEHKAHVTAHNPGVGHVALQGGSSSSPNAASLAAGKNGKMPRIHQP 60  
DB 1 melslgtehkaavhtaahnpvgvghvalqggsssspnaaslaaegknrgkmprihqp 60

QY 61 STAADGISAHQKSFSLRCIGTKPFSAPOGPTTHSKGATLRDLARDGGETQH 120  
DB 61 staadgisaaqhksfslrcigtkpfksapogptthskgatlrldlarddgetqn 120

QY 121 EAAAPDAARLTRSGVKKRRNDDMAGRMVKGSGGSEDKVPTQCKRHQLNNFGQMRTMLS 180  
DB 121 eaaapdaarltrsgvkkrrnddmagrmvkgsggedkvptqckrhqnnfgmrtmls 180

QY 181 KMAHPASANAGDRLOHSPPHIPGSHHEIKKEPPVGSFKATTAHADRVETAQEDDSEFOQ 240  
DB 181 kmahpasanagdrlohspphipgshheikeppvgstskattahadrvetaqedddsefqq 240

QY 241 LHQORLARENPPOPKLGAVATPSARFQPKLTAVAESVLGTDITQSPKLPQSMKGS 300  
DB 241 lhqorlarerenppopklgvatpsarfqpkltavaeavlgttditqspklpqsmkgs 300

QY 301 GAGVTPLATVLDKGLQAPNPPALNTLLKOTLGKDTQHYLAHHASSDGSQHLLDNKG 360  
DB 301 gagvtplatvldkglqapnppalntllkqtdqhylahhassdgsqhlldnkg 360

QY 361 HLFDIKSTATSYVLHNSHPGEIKLQAAGTGSVVDGSKGKISLGSGTQSHNKTMLSQ 420  
DB 361 hlfdikstatsysvlnshpgeiklqaagtgsvsvdgskgkislsgstqshnktmlsq 420

QY 421 PGEAHRSLTGTWOHPAGAARPOGSEIRLHDDKIHLLHPGLGWQSAKDXTDSQLSRQAD 480  
DB 421 pgeahrsltgtwohpagaarpgsesirldhdkihllhpelgwqsaadkxtdsqsrqad 480

QY 481 GKLYALKDNRTLQNLSDNKSSEKLYVDKTKSYSDVGQGVAILTDTTPGRHKMSIMPSLDAS 540  
DB 481 gklyalkdnrtlqnlstdnksseklvdktkysvvdgqgvailtdtppgrhkmsimpsldas 540

QY 541 PESHISLSLHFADAHQGLLHGKSELEAQSVATISHGRLVVADSEGKLFSAAIKPGQDGNEL 600

DB 541 peshislsfhfadahqglhlgkseleaqsvaishgrlvvadsegklfsaaikpgqdgnel 600

QY 601 KMKAMPQHALDEHFHDHQHISGFFHDDHQINLVKNNFROQHACPLGNDHQHPGWNLT 660

DB 601 kmkampqhaldehfhhdhqhsgffhddhqinalvknfnrqghacplgndhqhpqwnlt 660

QY 661 DALVIDNQLGLHHTNPEPHEILDWHLGSLALQBGKLIHYFDQLTGKWTGAESDCQKQKKG 720

DB 661 dalvidnqlglhhtnpepheildwghlsgslalqegklyhfdqltkwtgaesdcqklkg 720

QY 721 LDGAAYLLKXDEYKRLNINOSTSIKHGTENVFSLPHVRNKPPEGDALQGLNKKDQAQAM 780

DB 721 ldgaayllkxdevkrlningstssikhgtenvfslphvrnkpepdalqglndkdaqam 780

QY 781 AVIGVKNYKLTATEKDIRSQIKPGTQOLRPAQTLSREGISGELKDIHVHDKONLYALT 840

DB 781 avigvknlyaltetekdirsqikpgtqolrpaqtlsregisgelkdihvdknlyalt 840

QY 841 HEGEVFHPQREAWQNGAESSESWHKLALPQSESKLSKIDMSHEHKPIATFEDGSOHLKAG 900

DB 841 hegevfhpqreawngaessswhklaipqsesklksidmshehkpiatfedgshqklag 900

QY 901 GWHAYAAPRGPLAVGTSGSTQVFNRLMQGVKGVIPGSLTIVKLSAQGTGWTGAEGKV 960

DB 901 gwhayaaergplavgtsgstqvfnnrlmqgvkkgvipgslgtvklksaqgtgwtgaegrkv 960

QY 961 SSKFSRIRAYAFNPTMSTPRPKNAAYATQHGQWQREGKPLKPLYEMOGALIKQLDAHNV 1020

DB 961 skfserirayafnptmstprpknaayatqhgwqggtregikplyemogalikqldahnv 1020

QY 1021 HNAQPDLQSKLETLDLGEHGAELLNDKMRDELESATRSVTVLGHQHQVLSNGEIN 1080

DB 1021 hnaqpdlqskletldlgeghaellndmkrfdeleatsrsvtlvghqhvlsngein 1080

QY 1081 SEFKPSGKALVGSFVNNRSGDLSKSIQAVHATPPSAESKLSQSLGHLGHFVSAGVDMSHQ 1140

DB 1081 sefkpsgkalvgsfvnnrsgdlsksiqavhathpssaesklsglghghfvsagvdmshq 1140

QY 1141 KGEIPLGRQRPNDKATLTKSRLLDVTITGELHELADKAKLVSDHKPDQAIKOLRQOF 1200

DB 1141 kgeiplgrqrpndkaltksrllldvtitgelheladkaklvshdkpdqaiqlrqof 1200

QY 1201 DTUREKRYENPVKHYTDMGFTHNKALEANYDAVKAFINAFKKEHHGVNLTTRTVLESQ 1260

DB 1201 dtlurekryenpvkhytdmgfthnkaleanydavkafinafkkehgvnltrtrvlesq 1260

QY 1261 SAEALAKKLTNLLSLDSEGSMSFSRSGVSTVPTLSKKVPVPVPGAGITUDRAYN 1320

DB 1261 saelakkltllslsdgesmsfsrsgvgstvtptlskkvpvpvpgagitldrayn 1320

QY 1321 LSFRTSGGLNVSGRDRGVSGNIMVATGHDVMPYMTGKTKTSAGNASDWLSAKHKISPD 1380

DB 1321 lsfsrtsgglnvsgdrdgvsgnimvatghdvmptymtgkttksagnasdwlsakhkispd 1380

QY 1381 RIGAAVSGTLQGLQNSLKFKLTDELPGFTHGLTHTGTLTAEALLOKIEHOMKOGSKLT 1440

DB 1381 rigaavsgtlqgtlqnsllkfkltedelpgfthglthgtlttaellqkiehmqogsklt 1440

QY 1441 FSVDTSANLDRACINLNEGSKPNGVARTARVAGLSANLAAGSRERSTTSQGFSGTTS 1500

DB 1441 fsvdtsanldragnlnedgskpngvartarvsaglsasanlaagsersttsqgfsgtts 1500

QY 1501 ASNNRPFTLVNGVAGANLTAALGVASHSTHEGKPVGPFPATSTNVSAALADNRTSQSI 1560

DB 1501 asnnrpfvlngvgaganltaalgvashsthegkpvgifpattstnvsaaladnrtsgsi 1560

QY 1561 SLELKRAEPTVSNDSITSLTSLGHFKDSATTMLAALKELDDAKPAEQLHTLOOHESAK 1620

DB 1561 slelkraptvsnidsitstsltghfkdsattkmlaalkelddakpaeqlhltnoohesak 1620

QY 1621 DVVGDERYEAVRNKLKLVIRQOAAADSHSMELGSAHSTTYNNLSRINNDDGIVYELLKHF 1680

Db 1621 dvvgderyeavrnklk1virqaadshsmelgsashsttynnlsrinndgivellbkhfd 1680  
QY 1681 AALPASSARLGEEMNDPALKDIIKQLOSTPSSASVSMELKDGRLREOTEKATLDGKVG 1740  
Db 1681 aalpassakrlgemndpalkdiiqqlgstpfssasvsmelkdglreqtekailldgkvg 1740  
QY 1741 REEVGVLFODRNRLRVKVSQSVSKSEGFNTPALLLGTNSAAMSMERNIGTINFKYG 1800  
Db 1741 reevgvlfqdrnrlrvkvsqsvsksegnfcpalllgtnsaamsmernigtinikfy 1800  
QY 1801 QDQNTPRFTLEGGTAQANPOVASALTDLKEGLEMK 1838  
Db 1801 qdqntprftleggtaqanpqvasaltdlkeglemks 1838  
RESULT 2  
ID Y71095 standard; Protein: 1838 AA.  
XX AC Y71095;  
XX DT 08-SEP-2000 (first entry)  
XX DE Erwinia amylovora hypersensitive response elicitor encoded by dspE gene.  
XX KW Hypersensitive response elicitor; environmental stress resistance;  
XX KW plant; pathogen; dspE gene.  
XX OS Erwinia amylovora.  
XX PN WO200028055-A2.  
XX PD 18-MAY-2000.  
XX PF 04-NOV-1999; 99WO-US26039.  
XX PR 05-NOV-1998; 98US-0107243.  
XX PA (EDEN-) EDEN BIOSCIENCE CORP.  
XX PI Wei Z, Schading RL;  
XX WP: 2000-376566/32.  
XX DR N-PSDB; D00670.  
XX PT Application of a hypersensitive response elicitor protein to plants to  
XX PT impart stress resistance  
XX PS Disclosure; Page 15-20; 84pp; English.  
XX CC The patent discloses a method to impart stress resistance to plants by  
XX CC applying a hypersensitive response elicitor in a non-infectious form to  
XX CC a plant or seed. The present sequence is a hypersensitive response  
XX CC elicitor encoded by dspE gene from Erwinia amylovora.  
XX CC The protein elicits plant pathogen's hypersensitive response and is  
XX CC used to impart stress resistance to plants.  
XX SQ Sequence 1838 AA;  
Query Match 100.0%; Score 9448; DB 21; Length 1838;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1838; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MELKSLGTEHKAAYHTAAHNPVGHVALQOGSSSSPONAASLAEECKNRCMPRIHQ 60  
Db 1 melkslqtehaavhtaahnpvghvalqogsssspqnaaslaaegknrgkmprihq 60  
QY 61 STAADGISAHQKKSFLRGLCTKKFSRPAQOGQPTTHSKGATLRDLARDGETQH 120  
Db 61 staadgisaaahqkksflrgcltkkfsrpaqogqptthskgatlrldlardgetqh 120  
QY 121 EAAAPDAARLTRSGVGRNRNDDMAGRPWVGKGGSEDKVPTQQRHQLNPNFQMRQTMLS 180

Db 121 eaaapdaarltrsrgvgrnrnmdmagrpwvkgsgedkvyptqqrhqlnfnfgmrqtmls 180  
QY 181 KMAHPASANAGDRLOHSPPHIPGSHHEIKEEPPVGSSTKATTAHADRVETIAQEDDDSEFQQ 240  
Db 181 kmahpasanaagdrlohspphipgshheikeepvgstskattahadvrelaqeaddsefqg 240  
QY 241 LHQORLARERENPPOPPKLGVAATPISARPOPKLTAAVESVLEGTDTTQSPKPKQSMKLGKS 300  
Db 241 lhqorlarerenppppkpgvatpisarfqpkltavaesvleqtdtqspkpgsmkkg 300  
QY 301 GAGVTPPLAVTLDKGLQLAPDNPALNTLLKOTLQKDTQHYLAHHAASDGSOHLHLLDNKG 360  
Db 301 gagvtpplavtldkglqlapdnppalntllkqtdqthylahhassdgsqhlldnkg 360  
QY 361 HLFDIKSTATSYSVLHNSHPGEIKGLAQAGTGSVSDGKSGKISLGSSTQSHNKWTLSQ 420  
Db 361 hlfdikstatysvvlhnsphgeikglagagtgsvsdgskgislgstqshnktmlsq 420  
QY 421 PGEAHRSLLTGIIWOHPAGAARPOGESIRLHDDKIHILHPELGVWSADKDTSHLSROAD 480  
Db 421 pgeahrslltgiwqhpagaarppgesirlhddkihilhpelgvwsadkdtshlsrqad 480  
QY 481 GKLIALKDNRTLQNLSDNKSSEKLVDTIKSYSDVDRGQVAILTDTTPRRHKNSIMPDLAS 540  
Db 481 gklyalkdnrtlqnlndnksseklvdkiksydvdrqgvailtdtpprrhkmsimpsldas 540  
QY 541 PESHISLSLHPADAHQGLLHGKSELEAQSVAISHGRVLVVDSEGLFSAAIKPOGDGNEL 600  
Db 541 peshislslhfadahqglhlgkseleaqsvaishgrvlvvdseglfssaaiipkgdgnel 600  
QY 601 KMKAMPOHALDEHFQGHQISGFFHDDHGLNALYKNNFRQOHACPLGNDHOFHPQWNLT 660  
Db 601 kmkampqhaldhfqghqisgffhddhglqnalvknfrqhacplgndhfhpgwnlt 660  
QY 661 DALVIDNOLGLHHTNPEPHEITLDMCHLGLALOEKGLHYFDOLTKGWTGAESDCKQLKKG 720  
Db 661 dalvidnolglhhtnpepheitldmchlgslalqekghyfdoltkgwtgaesdckqlkkg 720  
QY 721 LDGAAYLLKDCGEVKRLNINQSTSSIKHGTENVFSLPHVRNKPPEPDALQGLNKDKAQAM 780  
Db 721 ldgaayllkdgevrkrlningstssikhgtenvfslphvrnkpepdalqglnkdkaqam 780  
QY 781 AVTCVNKYLALTEKGDTRSDQIKPGTQOLERPAOTLSREGISGELKDIHVDHKQNLAYLT 840  
Db 781 avigvnylaltekgdtrsdqikpgtqolerpaoqlsregisgelkdihvdhknqnlalt 840  
QY 841 HEGEVPHOPREAWONGAESSSWHKLALPQSESKLSLDMSHHEHKPIATFEDCSOHLKAG 900  
Db 841 hegevfhqpreawongaessswhklaipqsesklsldmshehkpiatfedgsgqhlkag 900  
QY 901 GWHAYAAPERGLAVGTSGQTVFNRLMQGVKGVIPGSLTIVKLSAOTGMTGAEGRRKV 960  
Db 901 gwhayaaperglavgtsgqtvfnrlmqgvkgvipgslitvkltsaotgmtgaegrrkv 960  
QY 961 SSKFSEIRIRAYANPTMSTPRPKNAAYATOHGQREGELKPLYEMQCALTKQLDAHNV 1020  
Db 961 sskfserirayantptmstprpknaayatohgqreglklplyemqcaltkqldahnv 1020  
QY 1021 HNAPODLOSKLETLDLGEGHAELLDNMDKFRDELQSAATRSVTVLGQHOGVLKNGEIN 1080  
Db 1021 hnapqdlqskletldlgeghaelldnmkfrdeleqsaatrsvtvlqhqgvkngeln 1080  
QY 1081 SEFKPSFGKALVSFNVNRSGDLSKSLQAAVHATPPSAESKLSMLGHFFSAGVDMSHQ 1140  
Db 1081 sefkpsfgkalvsfnvnrsgdlskslqqavhatppsaelkslsmghffsagvdmshq 1140  
QY 1141 KGEIPLGRORDPDNKTALTKSRLILDTVTIGELHELADKALVSDHKPDADQIQLROOF 1200  
Db 1141 kgeiplgrordpdnktaltksrilldvtvtigelheladkalvshdkpdadqiklroof 1200  
QY 1201 DTLREKRYESNPVKHYTDMGFTHNKALEANTDVKATINAFKKEHHGVNLTTRVLESOG 1260  
Db 1201 dtlrekryesnpvkhymtgftnkalleanvdvkatinafkehhgvnlttrvlesog 1260

1201	Db	dtlrekr	yesnpvkhytdmgfthnka	leanydavka	finafkkehghvnltrtvtlesgg	1260	
1261	Qy	SAELAKL	KNTLLSLD	SGESMSFRSYGGYSTVFVPTLSKKY	VPVPVPIPGAGITLDRAYN	1320	
1261	Db	saelakk	kntllsl	sgesmsfrsyggvstvfptlskkypvpvipgagitldrayn		1320	
1321	Qy	LSFSRTSGGLNV	SFRGDRGGVSGNIMVATGHDPMTGKTKTSAGNASDWLSAKHK	ISPDL		1380	
1321	Db	lsfsrtsg	lnvsgfrdgvgsgn	lmvatghdmpymtggktsagnasdwlsakhkispdl		1380	
1381	Qy	RIGAAVSGT	LQGTIONS	LKFKLTEDELPGFTHGLTHGLTPAEALLOKG	IEHOMKOGSKLT	1440	
1381	Db	rigaavsgt	lqgtlqns	lkfkltedelpgfthglthgltpaeallqkghmqkqsklt		1440	
1441	Qy	FSVDYSANL	DRAGTNLNEDGSKPNGVTRYASGLSASANAAGSR	SRTTSQGFSTTS		1500	
1441	Db	fsvdtsan	ldraglnlnedgskp	ngvtryasaglsasanaagsersttsqgfgstts		1500	
1501	Qy	ASNNRPTF	UNGVGAGANITAA	LGAVHSTHGBKPGVP	PPAPTSTNVSAALALNRRTSQSI	1560	
1501	Db	asnnrptf	ingvgaganitaa	lgaahssthegbkpgvipfapstcnvsaalalndrtsqsi		1560	
1561	Qy	SLELKR	AEVPTSDN	ISELTSTLGHKF	KFDSANTKMLAAULKELDDAKPAEQ	LHILHQHFSAK	1620
1561	Db	slelkr	aevptsdn	iseltstlghkfkdskatkm	laalldakpaeqhlilhqhfsak		1620
1621	Qy	DVVGDERY	EAVRNKLKLVIR	QQAADSHSMELGSASHSTTYNNLSR	INNDGIVELLHKHFD	1680	
1621	Db	dvvgdery	eaavrnlkklvir	qqaadshsmelgsashsttynnlsr	inndgivelhkhkhd		1680
1681	Qy	AALPASSAKR	LGE MNNDPALK	DIKIQLOSTPFSASVS	SMELKGLREOTFEKAILDQKVG	1740	
1681	Db	aalpassakr	lge mnndpalk	diiklqstpfssa	vsysselkgdretekailldkgv	1740	
1741	Qy	REEVGVL	FODRNNLRVKS	VSYSQSVSKSEGFNT	PALLGTSNSAAMS	MERNIGTFNKYG	1800
1741	Db	reevgvl	fodrnnlr	rvksvsqsvsksegfnt	pallgtsnsaams	mernigtinfkyg	1800
1801	Qy	QDQNTPRRTT	LEGGITAAQAN	PQVASALDILKKEGLEMS		1838	
1801	Db	qdqntprft	lenglaqan	pqvasalldilkkeglems		1838	

RESULT	3	
Y84856		
ID	Y84856	standard; Protein; 1838 AA.
XX		
AC	Y84856;	
XX		
DT	08-AUG-2000	(first entry)
XX		
DE	A hypersensitive response elicitor protein.	
XX		
KW	Hypersensitive response; insect control; disease resistance;	
KW	Hypersensitive response elicitor; plant growth; vegetable; crop;	
KW	ornamental plant; dspe gene.	
XX		
OS	Erwinia amylovora.	

XX	Hypersensitive response elicitor polypeptides useful for imparting
PT	enhanced growth, disease resistance and insect resistance to plants,
PT	especially vegetables and ornamental flowers -
XX	
PS	Disclosure; Page 17-22; 100pp; English.
XX	
CC	The present sequence represents a hypersensitive response elicitor
CC	polypeptide. The polynucleotide represents the dspe gene. The
CC	specification describes hypersensitive response elicitor polypeptide
CC	fragments, which do not elicit a hypersensitive response. Instead,
CC	the proteins impart disease resistance to plants, enhance plant
CC	growth, and/or control insects. The polypeptide fragments may be
CC	used to these properties to plants. The plants which may be treated
CC	in this way include vegetables, crops and ornamental plants such as
CC	alfalfa, rice, wheat, barley, rye, cotton, sunflower, peanut, corn,
CC	potato, sweet potato, bean, pea, chickory, lettuce, endive, cabbage,
CC	brussel sprout, beet, parsnip, turnip, cauliflower, broccoli, radish,
CC	spinach, onion, garlic, eggplant, pepper, celery, carrot, squash,
CC	pumpkin, zucchini, cucumber, apple, pear, melon, citrus, strawberry,
CC	grape, raspberry, pineapple, soybean, tobacco, tomato, sorghum or
CC	sugarcane, Arabidopsis thaliana, Saintpaulia, petunia, pelargonium,
CC	poinsettia, chrysanthemum, carnation or zinnia.
XX	
SQ	Sequence 1838 AA;
	Query Match 100.0%; Score 9448; DB 21; Length 1838;
	Best Local Similarity 100.0%; Pred. No. 0;
	Matches 1838; Conservative 0; Mismatches 0; Indels 0; Gaps
Qy	1 MELSLGTEHKAAVHTAAHNVPVGHVALOOGSSSSPONAAASLAEGKRGKMPRIHP 60
Ib	
Dd	1 melkslgtchkaavhtaahnpvgvhvalqggssssppnaaslaeagnkrgkmprihqp 60
Qy	61 STADGTSAAHQOKSFELRCIGTKFESRPAQSGPQTTHSKGATRLDLARDGETOH 12
Ib	
Dd	61 staadgtsaaahqqkksfslrclgcktkfsrapqgqqtchskgatrlldardgetqh 12
Qy	121 EAAAPDAARLTRSGGVKRNDMDMAGRPWVGSGGEDKVPTQOKRHQLNFGOMRTMLS 18
Ib	
Dd	121 eeaapdaarltrsggvkrndmdmagrpwmvkgsgedkvptqokrhqlnfgomrtmls 18
Qy	181 KMAHPASANAGRLOHSPHIPGSHHEIKEEPPVGSTSKATTAAHADRVETAQEDDSEFOQ 24
Ib	
Dd	181 kmahpasanagrloqshsphipgshheikeepvgstskattahadvetaaqeddsefq 24
Qy	241 LHQORLARERENPPPKLG VATPISARFPKLTA VAE SVLEGTDTTQSPLKPQSMKGS 30
Ib	
Dd	241 lhqqrларerenpppkpgvatpisarfqklitvae svlegtdtttqspkpqsmkgs 30
Qy	301 GAGVTP LAVTL DKG LQ LAPDNPPALNTLLKOTLGDTOHYLAHHASSDGSOHLLDNKG 36
Ib	
Dd	301 gavgtp lavtl dkg lql apdnppalntllkqlgkdtdhy lahhassdgsgqhlldnk 36
Qy	361 HLFDIKSTATSVVLHNSHPGEIKGLAQAGNGSVSDGSKGISUGSGTQSHNKTMLSQ 42
Ib	
Dd	361 hlfdlkstatsysvlhnsphgeikgl aagcgs vsvdgskgislgsgtqshnktmlsq 42
Qy	421 PGEAHRSLLTGTWQHPGAAR PQGESIRLHDCKIHLHP ELGWQ SADKDTHSQLRQAD 48
Ib	
Dd	421 pgeahrsltltgtwqhpgaarpqgesirlhddkihlhp elgwqs adkdthsqlrqad 48
Qy	481 GKLYALKONRTLONI.SDNKSSEKLVDKTKSYSDVRGQVAILTDPGRHKMSIMPSLDAS 54
Ib	
Dd	481 gklyalkdnrtlqnisdnsksekldvkiksyvsdvrgqvailtdtpgrhkmsimpsldas 54
Qy	541 PESHTLSLSHFADAHQGLLHGKSELFAQSVATSHGRLVVADSEGKLFSAAI PKQGDNEL 60
Ib	
Dd	541 peshls slsfhadahqgllhgk sel faqsvat shgrlvvadsegkl fsaai pkqgdnel 60
Qy	601 KNKAMPQHALDEHFHGDHOISGFFHDDHGO LNALYKNPNRQQRHACPLGNHDHOFPGWNLT 66
Ib	
Dd	601 knkamppo haldehfghdhoisgffhddhgo lnalyknpnrqqrhacplgnhdhohfpgwnlt 66



Db 601 kmkampqhaldehfgdhqisgffhddhggqlnaivknfrqhaqcaplgnqhfhpgwnlt 660  
QY 661 DALVIDNQLGLHTNPPEHLLDMGHLSALQEGKLIHYFDQLTKGWTGAESDCKQLKKG 720  
Db 661 dalvidnqlglhtnpphhlldmghlsglsalqegklyhyfdqltkgwtgaesdckqlkkg 720  
QY 721 LDGAAYLLDKGEVKRLNINOSTSIKHCTBNVPSLPHVRNKPPEPDALQGLNKDDKAQAM 780  
Db 721 ldgaayllkdgvekrlninostsihctbnvpslphvrnkppepdalqglnkddkaqam 780  
QY 781 AVIGVWKYIALTEKGIQIRSFQIKPGTOOLERPAOTLSREGISGELKDIHVDHKQNLXALT 840  
Db 781 avigvwnkylaltekgiqrfsfqikpgtqolrpaotlsregisgelkdihvdhknqnlxalt 840  
QY 841 HEGEVPHQPREAWQNGAESWHLKALPQSESKILSLDMSHEHKPIATPEDGSOHQKAG 900  
Db 841 hegevphqpreawngaeswhlkalpqseskilsldmshehkpiatpedgsohqkag 900  
QY 901 GWHYAAPERGPLAVGTSQSVFNRIMQGVKGKVIIPGSGLTVKLSAQTGSGMTCAEGRKV 960  
Db 901 gwhyaapergplavgtsgsvfnrlmqgvkgkvipgsgltvklsgaqtgsgmtgaegrkv 960  
QY 961 SSKFSRIRAYAFNPMTSPRPRIKNAAYATQHGQWREGKPLIYEMQCALIKQLDAHNR 1020  
Db 961 sskfserirayafnpmtsprpriknayatatqhgqwgregikplyemqcalikqldahnr 1020  
QY 1021 HNAPOPDLOSKLETLDLGHEGALLNDMKRFRDELSQATRSVTVLGQHGVLKSGEIN 1080  
Db 1021 hnapdpdlqskletldlgehaellndmkrfredeqsatrsvtvlghogvlsksgein 1080  
QY 1081 SEFKSPGKALVOSFNVNRSQDLSKLOQAVHATPPSAESKLSQMLGHPVSGVDMSHQ 1140  
Db 1081 sefkspgkalvgsfnvnrsqgdlkslqgavhatppsaesklsgmlghfvsagvdmshq 1140  
QY 1141 KGEIPLGRORDPNDKLTALTKSLILDTVTIGELHELDKAKLYSDHKPDADQIKLROOF 1200  
Db 1141 kgeiplgrordpndkltaltkslildtvtigelheladkalyshdkpdadqiklrqof 1200  
QY 1201 DTLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKKEHGVNLTTRTVLESQG 1260  
Db 1201 dtlrekryesnpvkhnytmdgftchnkaleanydavgafinafkkehgvnltrtrvlesqg 1260  
QY 1261 SAEALAKLKNTLLSDSGESMSFSRSYGGGVSTVFPVPTLSKKVPVPIPCAGITLDRAYN 1320  
Db 1261 saealakkntllsdsgeismfsrsygggvstvfptlskvpvpvpipcagitldravn 1320  
QY 1321 LSFRTSGCLNVSGRDGVSQINIVATGHDVMPYMTGKTSAGNASDMLSAKHKISPD 1380  
Db 1321 lsfrtsgclnvsgfrdgvsqinivatghdvmptgktsagnasdmlsakhkispd 1380  
QY 1381 RIGAAVSTLOGTLONSLKFKLTDELPGFIHGLTHGTLTPAELLQKGIHQMKQSGKIT 1440  
Db 1381 rigaavstlogtloonslkfkltdelpgfihgltghgtltpaellqkgihqmkqsgkit 1440  
QY 1441 FSVDTSANLDLRAGLNLEDGSKNGVTARVSAGLSASANLAAGSPRSTTSQFGSTWS 1500  
Db 1441 fsvdtanldlragnlnledgskngvtarvsaglsasanlaagsprsttsqfgstws 1500  
QY 1501 ASNNRPTFLNGVGAGANLTAALGVAHSSTHEGKPVGIFPAFTSTNVSAALADNRSTSQT 1560  
Db 1501 asnnrptflngvgaganltaalgvassthegkpvgiifpafstnvsaaladnrstsqst 1560  
QY 1561 SLELKRAEPTVNSIDELSTSLGKFKDSATTKMLAALKELDDAKPAEQLHILQOHFSAK 1620  
Db 1561 slelkraptvnsidelslsgkfkdsattkmlaalkelddakpaeqhlilqhfsak 1620  
QY 1621 DVVGDERYEAARNLKLVIROQADSHSMELGSGASHSTYNNLSRINDGIVELLKHKFD 1680  
Db 1621 dvvgderyearnllklviroqadshsmelgsgashstynnlsrindgivelhkhkfd 1680  
QY 1681 AALPASSAKRLGEMMNDPALDKIIKQLQSTPFSASYSMELDKGLREQTEKAILDGKVG 1740  
Db 1681 aalpassakrlgemmnndpaldkiiqqlqstpfssasysmelkdglreqtekailldgkv 1740

QY 1741 REEVGVLFODRNNLRVKSVSVSKSEGFENTPALLLGTNSAAMSWERNIGTINPKYG 1800  
Db 1741 reevgvlfodrnlnrvksvsvsksegfncpalllgtcnsaamsmwnigntinkyg 1800  
QY 1801 QDQNTPRRTLEGGIAOANPQVASALTDLKKEGLEMKS 1838  
Db 1801 qdqntrprfleggiaanpqvasaltdlkkeglemks 1838

## RESULT 4

Y69258

ID Y69258 standard; Protein; 201 AA.

XX Y69258;

XX AC

XX 30-MAY-2000 (first entry)

XX DT

XX Amino acid sequence of the secretion signal of the DspE protein.

XX Type III secretion system; DspE protein; hairpin secretion system;

KW effector protein; Avr protein; avirulence protein; agriculture.

XX Erwinia amylovora.

XX WO200002996-A2.

XX 20-JAN-2000.

XX 09-JUL-1999; 99WO-US15425.

XX 10-JUL-1998; 98US-0092357.

XX (CORR ) CORNELL RES FOUND INC.

XX Bauer DW, Beer SV, Bogdanove AJ, Collmer A, Ham JH;

XX WPI; 2000-182218/16.

XX New DNA construct encoding type III secretion system, used for

PT recombinant production of secreted protein and for identifying

PT potential effector proteins -

XX Claim 6; Page 14-15; 71pp; English.

XX The present sequence represents a functional type III secretion

CC system from the N-terminal domain of the DspE protein. The DspE

CC protein is secreted by a type III secretion system. The DspE secretion

CC signal is compatible with the hairpin secretion system of Erwinia

CC amylovora. The secretion signal polynucleotide is used to produce

CC the constructs of the invention, which also comprise a promoter and a

CC second DNA that encodes a protein or polypeptide that can be secreted

CC by the type III secretion signal. Host cells containing the constructs

CC are used for recombinant production of the polypeptide. The constructs

CC are also used to screen for potential effector proteins, e.g. Avr

CC (avirulence) proteins potentially useful in agriculture.

XX Sequence 201 AA;

## Query Match

Best Local Similarity 11.0%; Score 1041; DB 21; Length 201;

Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MELKSLCTEHAHVTAHNPVGHGVALQQSSSSSSPONAAASLAEEKNGKMPRIHQP 60  
Db 1 melkslctehaavhtaahnpvghvalqgssssspnaaaslaaekngkmprihqp 60  
QY 61 STAADGISAAHQKKSFLRCGLGTTKFFSAPOGPCTTHSKGATLRDLARDGGETQH 120  
Db 61 staadgisaaahqkksflrcglgttkffsapgpgtthskgatlrldlarddgetqh 120  
QY 121 EAAAPDAARLTRSGGKRRNRNDDMAGRPVYKGGSGEDKVPVQQKRHQLNNFGOMHQTMLS 180

Matches	389;	Conservative	287;	Mismatches	803;	Indels	586;	Gaps	91;
QY	7	GTH- <b>KA</b> AVHTAAHPV- <b>---</b>	GHGVALQOGSS- <b>---</b>	SSPQNAAS	43				
Db	37	gsahvsvpftgthapvcrsnifslglsclavgtaniafadgiadkaapktqat	96						
QY	44	LA <del>A</del> EGNKGKMPRIHOPSTAA <del>D</del> GISAAHOO <del>K</del> SFSURGLCTKKFSR <del>S</del> APOGPG- <b>---</b>	98						
Db	97	ilqtgng- <b>---</b> lpqvnigtptsgvsnvqafvgmrgailnn- <b>---</b> srstnqtqlggwiqg	151						
QY	99	<b>---</b> TT <del>H</del> SK <b>---</b>	GATLRDLLARDG <b>ET</b> OH <del>E</del> AAAA <del>P</del> AARL <b>TR</b>	132					
Db	152	npwlargearvvvnqinshssqmgnyievgrraeavvianp <del>gi</del> avngggffinasatl	211						
QY	133	SGGKRRMDDMAGRPVKG <b>---</b>	GSGED <b>---</b>	KVPTQKRHOLNFGQMOPMLS	180				
Db	212	ttgqpyqagdisgfkirqgnvviaghgdardtdfrilsyhs <del>k</del> idapwggdvr <del>v</del> ag	271						
QY	181	KMAHPASANAGDRIQHSGPHIPGSHHEIKKEPVGSTSKATAHADRVETIAOEDDDSEFQQ	240						
Db	272	qndvva <b>tg</b> na <b>---</b> hsp <b>---</b>	312						
QY	241	LHQQLARERENPPQKLG <b>VAT</b> ISARFOPKLTAVAESVLEGTDTTQSPKLQPSMLKGS	300						
Db	313	<b>---</b> klg <b>---</b>	gmyankitlista <b>---</b>	egaglrnqqqlfas	342				
QY	301	GAGVPLAVTLDKGLQIAPDNPALNTLLKQTLGKTOTQHYLAHHAS <b>---</b>	SDGSQHLLLDN	358					
Db	343	sgn <b>---</b> validangrlvnsqgmaaa <b>---</b>	aktdntaehkvnirsgg <b>---</b>	ven	385				
QY	359	KG <b>---</b>	HLFDIKSTATSYS <b>---</b>	VLHNSHPGEIKGLAOGTGSVSDGKSGKI	404				
Db	386	sgtavsggtgihsqslqngtclllssgcilhms <b>---</b>	gslknetsgti <b>---</b>	eaarl	435				
QY	405	SLGSOTOSHNTML <del>S</del> OPCEAHRSLTGTI <del>Q</del> WHPAGAAPRGESIRLHDDKIHILHPELG <b>VW</b>	464						
Db	436	aidtdt-lnnqgklsq <b>g</b> <b>---</b>	<b>---</b>	sgklhi <b>---</b>	458				
QY	465	QSAEDKTHSQLSRQADGKLYALKONRTLQNLSD <b>---</b>	NKSSKLVKDVKISYSDVRGQ	518					
Db	459	<b>---</b> daqqkm <b>---</b>	dnrgmgqlcptasngsnqtgn <b>---</b>	synasfhss	498				
QY	519	VAILTDTGRHKMSIMP <del>S</del> LDASPEHSISLSLHFADAHQGLHGKSELEQAQSV <del>A</del> ISHGR <b>LV</b>	578						
Db	499	ttpttaigtgtatvsnitapt <b>---</b>	fad <b>---</b>	gtirthgalnsgsiiangtd	547				
QY	579	VADSEKFLFSAAIPKQGGCNELKMKAMPQH <del>A</del> LDEHFGHDHOISGFFHDDHQLNALVKN <b>---</b>	637						
Db	548	vsaggg-lonnag <b>---</b> qlidhqlnag <b>---</b>	safdnhng <b>---</b>	tlisдавhiqagslnnqngni	598				
QY	638	NFRQHCAPLGNDHQFHQPGWNLTDALVID <b>---</b>	NOLGLHHTNPPEHPEILDDMGHLGSLALQ <b>EG</b>	696					
Db	599	tttrqg <b>---</b>	leletqldnahgkllisaeiadlavsgslnnqng	638					
QY	697	LHYFDQLTKGWTGAESDCQKKGKLDGAA <del>Y</del> LKDG <del>E</del> VRRLMINOSTSIKKGTE <b>NV</b> SLP	756						
Db	639	iatnqql <b>---</b>	lihdqqgstavidntngtligsdv <del>ai</del> qak	675					
QY	757	HVRNPEPEDALQGLNK <b>---</b>	DDKAQAMAVIGYKNVLYALTEKDGIRSFQIKPGCTQ <b>OLE</b>	810					
Db	676	slsn <b>---</b>	ngtlaadnkdialqldofyvernliavagne-lsistrqslknshtlqagk <b>ir</b>	730					
QY	811	RPAQTLSREG <b>---</b>	ISGELKDIHVDHQNLYALTHEGEVFFHOPREAWONGAES <b>SW</b> HKLA	866					
Db	731	ikannldnaaagniqsggttdigtqhn <b>---</b>	ltnrgliddg <b>---</b>	767					
QY	867	LPQSESKLSLDM <b>SH</b> <b>---</b>	EHKPTATFEDGSGHQLKAGGHAYAAPERGLAVG <b>TS</b>	918					
Db	768	<b>---</b> ktkiqagqmni <del>g</del> triygdnaiaatrlndngt <b>g</b> a <b>---</b>	aiaaenlnlgig <b>---</b>	819					
QY	919	GSOTVFNRLMQGVKGKVP <del>IG</del> SLTVKL <del>S</del> AQTGGMTGAEGKRVSSKFSERIRAF <b>NPT</b> MS	978						
Db	820	<b>---</b> qlnrepslsvngndmavagaldtnogatkag <b>ir</b> <b>---</b>	<b>---</b>	853					

Db	121	eaapdaarltrsgvgvkrnmddmagrvmvk9gsgedkvtpcqqrhqlnfnfgmrqtmls	180
Qy	181	KMAHPASANAGDRLQHSPPHI	201
Db	181	kmahpasanagdrqlchspphi	201
RESULT	5		
ID	Y75097		
XX	Y75097 standard; Protein; 2514 AA.		
XX	Y75097;		
DT	21-MAR-2000 (first entry)		
DE	Neisseria meningitidis ORF 564 protein sequence SEQ ID NO:1668.		
KW	Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;		
KW	antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;		
KW	antibacterial; gene therapy.		
OS	Neisseria meningitidis.		
XX	WO9957280-A2.		
PD	11-NOV-1999.		
XX	30-APR-1999; 99WO-US09346.		
PR	01-MAY-1998; 98US-0083758.		
PR	31-JUL-1998; 98US-0094869.		
PR	02-SEP-1998; 98US-0098994.		
PR	02-SEP-1998; 98US-0099062.		
PR	09-OCT-1998; 98US-0103749.		
PR	09-OCT-1998; 98US-0103794.		
PR	09-OCT-1998; 98US-0103796.		
PR	23-FEB-1999; 99US-0121528.		
XX	(CHIR ) CHIRON CORP.		
PA	(GENO-) INST GENOMIC RES.		
XX	Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;		
PI	Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;		
DI	Tettelin H, Venter JC;		
XX	WPI; 2000-062150/05.		
DR	N-PSDB; Z53859.		
XX	Novel Neisserial polypeptides predicted to be useful antigens for		
PT	vaccines and diagnostics		
PS	Claim 2; Page 854-855; 1453pp; English.		
CC	Z53015 to Z54536, Z54577 to Z54615, and Y74253 to Y75941 represent		
CC	novel Neisseria meningitis and N. gonorrhoeae polynucleotides and		
CC	polypeptides. Z54537 to Z54576 and Z54616 to Z5473 represent PCR		
CC	primers used in the exemplification of the present invention. The		
CC	polypeptides, the polynucleotides, antibodies and compositions of		
CC	the invention can be used as vaccines as diagnostic reagents, and as		
CC	immunogenic compositions. The polypeptides can be used in the		
CC	manufacture of medicaments for treating or preventing infection due to		
CC	Neisserial bacteria (e.g. meningitis and septicemia), to detect the		
CC	presence of Neisseria bacteria, or to raise antibodies. They may also		
CC	be used to screen for agonists or antagonists, which may themselves		
CC	have use as antibacterial agents. The polynucleotides of the invention		
CC	may also be used in gene therapy protocols.		
XX	Sequence 2514 AA;		
SQ			
Query Match	2.9%; Score 275; DB 21; Length 2514;		
Best Local Similarity	18.8%; Pred. No. 3.6e-08;		



Db 687 tvn-----iqsqtlngsqhateqitinsrvndnngkllsanqaglavsdglyndhge 742  
Qy 157 DKVPTQQRKHQIINNCOMQOTMLSKMAHAPASANAGORLOHSPHIPGSHHEIKEEPVGST 216  
Db 743 latnrglsihdkn-----qntialnmdagtiqagn-----vslqaksla 782  
Qy 217 SKATTAHADVEIAOEDDSEFOOLHQORLAREFENPPPKLGVAIPISARFQPKLTAV 276  
Db 783 nngtltagnkidlaltd-----fvverd-----ltagqlnls-- 816  
Qy 277 AESVLEGTDTTOSPLKQSMKSGAGVTPLAVTLDKGKLQLOAPDNPALMTLLKQTLGK 336  
Db 817 ikgrlnkthtlq-----aght-----lknagni-----dn-----qvtgk 847  
Qy 337 DTQHYLAHHASSDGSQHLLOLNDKGLHFDIKSTATSYSVLHNSHPGETKGLQAAGTGSVS 396  
Db 848 -----ilggeqtditseqhvdnrg-----lins-----dglthigagqtl 882  
Qy 397 VDGKSGKISLGSGTQSHNKTMLSGPFAHRSLLTGIWOHPAGAARPOGE--SIRLHDDKI 454  
Db 883 tntgtoki-ygnhialdaqillnreetegstkag-----aiaarkridigakeihndeg 936  
Qy 455 HILHPE--LGWQSDAKDTHSQ-----LSRQADGKLVALKDNRTLQNLSDNKS 500  
Db 937 alissegifavgnrdeqhaagmadfvngsaglevggd-----almsvrmmqnnhfrk 992  
Qy 501 SE----KLVDKIKSYV-----DQRQVAILTDPGRHKMSIMPSLDASPESHISLS-L 549  
Db 993 tetylakaekqvdytvgntyyqagkdglfdnsqgq-kdqtatfhlkngsrieangw 1051  
Qy 550 HFADAH-----QGLLH-----GKSELEAQSVASHGRLLVAD-----581  
Db 1052 hvrdyhietkerlienpahitvvgdltasgqwnkdsrivrvggriitddlnqkeith 1111  
Qy 582 --SEGKLFSAAIKPGQD-----GNELMKAMPOHALDEBHFHGHQISGFHDDHGO 630  
Db 1112 qsttgkrtavgtqwdsvtkkwygsgkrqriternhtp-----yhdql--fthdftp 1165  
Qy 631 LNALVKN-----NFRQOQHAC-----PLGNDHQFHPC--WNLTDALVIDNQLGLHHTN 675  
Db 1166 vsvliqnaaspsfpapasailldgvtasvnggrihtgnvvvlnnatvtlplnslyth 1225  
Qy 676 PEPHEILDGMHGLSLAQEKLHVPDLTKGWTGAESDCOLKGLGDGAAYLLKDGSEVKR 735  
Db 1226 pd-----nkgw-----lvtdpqfady--rrwlgdymllqqiqltdnlnhkrldgdyve 1273  
Qy 736 LNIQSTSSIKHGTENFSLPHVRNKPDPGDALQGLNKKDQAQAMAVIGNVKYLALTEK 795  
Db 1274 klvn-----eqibqltgyrr-----ldgyrsdee-----gfkalmndg 1306  
Qy 796 --DIRSFOIKPC--TOQ-----LERPAOTLSREGISGELKDIHVDHKKQLLYAL 839  
Db 1307 ltaaktfgltppialsaeqvarltsdivmenqvtls-----dgsqtvlv--pkvyal 1359  
Qy 840 THEGEV-----PHQPREAWQNAESS--WHKLALPOSES-----872  
Db 1360 arkgdlnstsgllisaevllkqlngnltnsgtiagrqavliqarninsngniqadqiglk 1419  
Qy 873 KLKSLDMSEHKPIATFEDGSQH-----LKAGGWHAYAAAPER-----G 911  
Db 1420 aeksinl-----dggvqaggrllltaqaqnlnlntgttsgrnngntaidrmag 1468  
Qy 912 PLAVGTSGSQTVFNRLMOGVKVIIPGSGITVKLSAQGTGOMTGAEGRKVSCKFSE----RI 968  
Db 1469 invvg-shteqvndrtsdgi-----slhasndinlnaatvsngvkdgttqi 1514  
Qy 969 RAYAFNPMTSTPRPKNAAYATQHWQHGREGLEKPLVEMQCALIKQLDHAHVRNNAPODL 1028  
Db 1515 ta-gnnlnlgtirtetareygt-----lodenhrhvrqrstev 1550  
Qy 1029 OSKLETLDLGEHGAEL--NDMKRFRDELFPQ-----SATRSVT-----VL 1066

Db 1551 gssirt---qngallragndklirgqealeeagktvlaagrdvtisegrqriteidsvs 1606  
Qy 1067 GOHQGVKLSNGEINSEFKPS-----PKALV---QSFNVN-----RSG 1101  
Db 1607 gkskgilssst-kthdryrfshdeavgsnigggkmiavaagddinvrgsnliskdgivlkag 1665  
Qy 1102 QDLKSLSLQQAQVHATPPSAESKLSQMLGHFVAGVDMSHQKGEIPLGROR--DPNDKTALT 1159  
Db 1666 hdidistahrytqneyheskksygm--tggigft-----ignrkttdtdtrniv 1715  
Qy 1160 KSRLLI-----DVTIIGELHELADAKLIVSDHKPDADQIKQLRQOFDTLREKRYESNPVK 1214  
Db 1716 htgsiilngldgtvtvagnryrqtgstvsspegrntvtaksidvef---annriyaday 1772  
Qy 1215 HYTDMGFT---HNKALEANYDAVKAFINAFKKEHGVNLTTRTVLESQCSAEALAKKLKNT 1271  
Db 1773 tteqkgitvalnvppvvaagqnfiaaagvngksknkrvmaanaaawq-syqatqmqmf 1831  
Qy 1272 LLSLDGSEMSFSRSYGGGVSTVFVPTLSK-----KVPVPVIPCAGITLDRATNLSF 1323  
Db 1832 apssaaaggggnnqpsisvitygeqksrneqkhyteaasqilgkgt-----t1 1884  
Qy 1324 SRTSGG---LNVSEGRDGVSGNIMVATGHDVM--PYMTGKTSAGNASDW--LSAKHKI 1376  
Db 1885 aatgsgeqsninitgsdvighagtalladnhrisakdgseqsknksngwagvavki 1944  
Qy 1377 SPDRLIRAAVSGTL--QGTQLNSLKFTEDELPGFIHGLTH-----CTLT 1420  
Db 1945 gngirfgitagnigkgeqgg-----stchrthvgsttgkttirsggdttk 1993  
Qy 1421 PAELLQKIE---HQMKG-----GSKLT---1440  
Db 1994 gvqilgkigadtrnlhiesvqdtetygskqngnqvvtvgyfsaagsyrqskvadh 2053  
Qy 1441 -----FSDVTSAANLDRAGINLNEDEGSKPNGVTVARVSAGLSASANLAGSR 1486  
Db 2054 svtgsggiyagedgyikvrndtdlkggilitssqsaedkgknlftatlltas-diqnhr 2112  
Qy 1487 ERSTTSOGFS-----TTSASNNRTFFLNGVAG-----ANLTAALGV---AH 1526  
Db 2113 yegrsfjggsfdlmgwdgtvtdkqgrptdrispaagysgdgsknsttrsgvntnih 2172  
Qy 1527 SSTEKGKPVGIPAFPTSTNVSAALALDNRTSQSISLELKRA--EPVTSNDIS---ELTST 1581  
Db 2173 itdeagqlartortaketeariygtidtetadqhsghlknksfdkdavakeinlqrevtke 2232  
Qy 1582 LGKHFKDSATTKMLAALKELDDAKPAEQHLILOQHFSACKDVGVGDVEAVNKLKLVIRQ 1641  
Db 2233 fgrn-----aaqavaav-----adklqntdsy-----eryqearfl--leael 2268  
Qy 1642 QAADSHSME-----LGSASHSTTYNNLSRINN 1668  
Db 2269 qntdseaeakaafasrlgqvnaaylaenqsyrydtwkeggigrsilbgaaglltgsjglla 2328  
Qy 1669 DGIVELLHKKHFDAAALPASSAKRILGEMMNNDPALDKIIKLOQSTPFSSASVSMELKDGLRE 1728  
Db 2329 gggtslaaqyldka-----aenlg-----pagkaavnaalgaagaigayt-----2366  
Qy 1729 QTEKAILDKGVGEEVGVLPQDRNN--LRVKSVSVS-----QSVSKSEGFNTPALIL 1778  
Db 2367 -----gsgsgavvganv-dwnnrqlhpkemaladyaealkrevkregrki-----2412  
Qy 1779 GTSNSAAMSERNI 1792  
Db 2413 -ssqeaamrirqi 2425

RESULT 7  
Y85564  
ID Y85564 standard; Protein; 1837 AA.  
XX Y85564;  
XX





Db 558 dnaahitgtskltaskdmvgveagllnvntntrntnrtsgnllhicaakgni-----qlrntkl 613

Qy 761 ---KPEPGDALOGLNKDDKAQAMAVIGVNYKLA-----LPEKGDIRSFQIKPGT 806

Db 614 naakalctaldgnlvdsdglhavsadghvslslangnadftghntltakadnvnagsvgkgr 673

Qy 807 QOLERPAOTLS-----REGISG-----ELKDIHVDHKQ- 834

Db 674 ikadntntssgdltlvagnglqldgkgrnsingkhisiknngnaddklninvhaksg 733

Qy 835 ---NLFA-----LTHEGEVPHQPREAWONAE-----SSSHKLLALPQ 869

Db 734 alnihsdralstientklesthlnaqlhervtlnqvdavahchlsitgsiqwqndklp- 792

Qy 870 SESKLKS---LMSSEHKPIATFEDGSQHOLKAGGHWAAAP---ERGPLAVGTSSQTV 923

Db 793 sanklvagvlnarysgia-----dntlragealnltagtalvkrgrlnwstvatktl 847

Qy 924 -FNRLMQGVKKG--VIPGSG-LTV-----KLSAQT-----GGMTGAEGRKV 960

Db 848 ednaelkplagrlnieagsgtltiepanrisahtdlsiktggklllsakggnagapsaqv 907

Qy 961 SS---KFSEIRIYAFNPTMTPRTPKNAAYATOHGWOGREGUKPLYEMOGALIKOLDIAH 1017

Db 908 ssleakgnrlvtgetdlrgskitagknlvattkgnlieavnnsfsnyfptqkaael- 966

Qy 1018 NVRHNAPODLOSLETLDLGEHGAELNDMKFRFRDELE---QSATRSVTVLGHOHVLR 1074

Db 967 ---nqkskeleqiaqlksspskklptlqgeerdlafyiqainkev- -gk-----k 1015

Qy 1075 SNGEINSEFPKSPGKA-LVQSPNVNRSQDLSKSLQAVHAT---PPSAESKLSQMLGHF 1130

Db 1016 pkgkeylqaklsaqnidlisagieisgdsitaskklnlhaagvlpkaadseaaail--- 1072

Qy 1131 VSAGVDMSHORGEIPLGRORDPNDKLTATK-SRLILDVTIG-ELHELA---DKAKLVSDH 1186

Db 1073 ---ldgtdgdeigkptykshykaalnksprl---tgrtgvsihaaalddarilii-- 1123

Qy 1187 KPDQDIKOLROOFTLREKRYESNPVKHYTDMGFTHNKALPANYDAVKAFINAFKKEHH 1246

Db 1124 -gaseikapsqid-----ikahsdi-----vleagqndayflltkgksk 1164

Qy 1247 GVNLTTRT-----VLESQGSAEFLAKKLKANT---LLSLDGSGESMSF- 1283

Db 1165 lirktkftstrdhlmpapveltaangitlqaggnieanttrfnapagkvtlvageelqll 1224

Qy 1284 -----SRSYGG---GVSTVEVPTLSK-KVPVPVTPGAGITLDRAINLSFSR 1325

Db 1225 aeegihkhelddvqksrrfrikvgksnyskneinetklpvrvaqtaatr----- 1274

Qy 1326 TSGGLNVSGRDGVSNGNIWATGHDVMPYMTGKKTTSAG-----NASDWLS 1371

Db 1275 -----sgwdtvltegtfkttiagadiqagvgekarvdakiilkgivnrq 1319

Qy 1372 AKHKISPDILRGAAVSGTLQGLQNSLPKFLFEDEL- -GFIHGLTHGTL-TPAE 1423

Db 1320 seekletnstvwqkqag--rgstielklpsfesptppklsapggviyvdipkgnlktele 1377

Qy 1424 LLOGGIEH-QMKOGSKLTFSDVTSANLDRAGINLED-----GSKPNGVTAR----- 1470

Db 1378 klskapeyaylkq-----lvaknlnwqvlqaydrwdykqeglteagaail 1424

Qy 1471 -----VSAGLSASANLAAGSRERSTTSQFGSTTSASNNRPTFLNGVAGANLTAALGV 1524

Db 1425 alavtvttsagtgavlglnaataadaafasl--asqasvfinnkdvgtktikelig- 1481

Qy 1525 AHSSTHEKPGVGPAPFTSTNVSAALDNRTSQ-----SISLELKRAEPTVNDISELTS 1580

Db 1482 -rsstvtknlvvaatagvadvkligas-alnnvsvdkqwnlnltvnlanaagaalinn---ca 1535

Qy 1581 TLGKHFKDSATTKMLAALKELDDAKPAQOLHILQOHFSKADV----- 1622

Db 1536 vngslkdnleannlilaalvntahgeaaskikldqnylvhkhahaiagcaaaaankgkq 1595

Qy 1623 -----VGDERYEAVRNLLK-----LVIROQAADSHSMELGSASHSTTYNNL 1663

Db 1596 dgaigaavgealtnvgealtnknpdtltakereqilaysklvagtvgvgdvnaanaae 1655

Qy 1664 SRINDGIVELLHKHFDAAALPASSAKRLGEMMNDPAL--KDIKLOLOSTPFSASVSME 1721

Db 1656 vavknqldskregrefndemtacqg-----nnpqlcrkntvkkkyqnvadvkrllaasia 1708

Qy 1722 LKDGIREOTE-----KALDGKV--GREEVGVLFODRNNLRVKSVSQS--VSKSEGF 1771

Db 1709 ictdisrteclrtirkqlidrsihssweagligkddewyklfskysqtadialqsyhl 1768

Qy 1772 NTPA-LLLGTSNSAAMS-----MERNIGTINFKYGO----- 1801

Db 1769 ntaakswlqsgntkplsewmsdqqytlisgvnprfipgrgvkqntptnkvypegisf 1828

Qy 1802 DQNTPRR-----FTLEGGIAQANPQVASALTDLKKKEGLEMKS 1838

Db 1829 dtnlkrhlanadgfsqeqgikgahr--tnfmaelnsggrvks 1870

RESULT 9

Y85574 ID Y85574 standard; Protein; 2608 AA.

XX AC Y85574;

XX DT 07-JUL-2000 (first entry)

XX HS-UNC-53/3/GFP fusion insert of plasmid pG13303.

XX DE UNC-53; Caenorhabditis elegans; microtubule; neural regeneration;

XX KW anticancer; anti-neurodegeneration; antifibrotic; anti-adhesive; human;

XX KW antisclerotic; antimetastatic; anti-arthritis; autoimmune disease.

XX OS Synthetic.

XX OS Homo sapiens.

XX PH Key Location/Qualifiers

XX FT Misc-difference 1194 /note= "unspecified"

XX WO963080-A1.

XX PN 09-DEC-1999.

XX PD 02-JUN-1999; 99WO-EP03848.

XX PF 03-JUN-1998; 98GB-0011962.

XX PR (JANC ) JANSSEN PHARM NV.

XX PA Luyten WHML, De Raeymaeker MC, Geysen JJGH, Bogaert TAOE;

XX PI Maerten LJS, Verhasselt P, Van De Craen M;

XX XX WPI; 2000-116370/10.

XX DR N-PSDB; A07846.

XX PT Novel proteins and nucleic acids e.g. for treating neurodegeneration -

XX PS Disclosure; Fig 7e; 146pp; English.

XX CC The invention provides vertebrate (human) protein homologue of a UNC-53

XX CC protein of Caenorhabditis elegans. The UNC-53 binds to microtubules or

XX CC their plus ends. The UNC-53 sequences are used to promote neural

XX CC regeneration, revascularization and wound healing; also for treating

XX CC neurodegenerative disease, acute traumatic injury, fibrotic disease and

XX CC autoimmune diseases (e.g. rheumatoid arthritis and sclerosis). The UNC-53

XX CC polynucleotides can be used for recombinant production of the proteins,

XX CC as a source of probes for detecting allelic variants and polymorphisms,

XX CC for sequencing genomic DNA and for detecting UNC-53 expression; and as

XX CC source of therapeutic antisense sequences. Cells that express the





Db 2020 aeilqlkselrekelkltdirlea-----lssahhldqireamrmqne--ieilka 2070  
QY 1678 HFD-----AALP-----ASSAKRLGEMMN-----DPALKDII-----1705  
Db 2071 endrlkaetgnaktptrppsesssstssrsqslglsinnlniteavssdilldagda 2130  
QY 1706 -----KOLQSTPFSASVSMELKQGLRQTEKAILDGKVR--1741  
Db 2131 tghkdgrsvkiisvsgygrakdqkqayllgsi-----gvsgktkwldvgvirlf 2184  
QY 1742 EEVGVLFQDNNRLRVKSVSVSQ-----SVSKSEGFNTPALLLGTSNSAAMSERNIGTINF 1797  
Db 2185 keyvfridtslglssdciasycigdlirshnlvpeell-----pcgylvgdnniitvnl 2240  
QY 1798 KYGODQNTPRRFTLE 1812  
Db 2241 k-gveensldsfvfd 2254  
  
RESULT 10  
B41087  
ID B41087 standard; Protein; 2541 AA.  
XX  
AC B41087;  
XX  
DT 08-FEB-2001 (first entry)  
XX  
DE Human ORFX ORF851 polypeptide sequence SEQ ID NO:1702.  
XX  
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
KW vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
KW hypotensive; dermatological; immunosuppressive; antinflammatory;  
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;  
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
KW cholesterol ester storage; systemic lupus erythematosus; infection;  
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
KW bone damage; cartilage damage; antinflammatory disease; coagulation;  
KW thrombosis; contraceptive.  
XX  
OS Homo sapiens.  
XX  
PN WO200058473-A2.  
XX  
PD 05-OCT-2000.  
XX  
PF 31-MAR-2000; 2000WO-US08621.  
XX  
PR 31-MAR-1999; 99US-0127607.  
PR 02-APR-1999; 99US-0127636.  
PR 05-APR-1999; 99US-0127728.  
PR 30-MAR-2000; 2000US-0540763.  
XX  
PA (CURA-) CURAGEN CORP.  
XX  
PI Shinkets RA, Leach M;  
XX  
PT WPT: 2000-602362/57.  
DR N-PSDB; C75296.  
XX  
PT Novel nucleic acids and peptides derived from open reading frame x,  
PT useful for treating e.g. cancers, proliferative disorders,  
PT neurodegenerative disorders and cardiovascular disease -  
XX  
PS Claim 11; Page 1357; 5507pp; English.  
XX  
CC C74446 to C77606 encode the proteins given in B40237 to B43397, which  
CC represent the human ORFX open reading frames 1 to 3161. The ORFX  
CC sequences have activities such as: cytostatic; hepatotropic; vulnery;

CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective; osteopathic;  
CC anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;  
CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;  
CC dermatological; immunosuppressive; antinflammatory; antibacterial;  
CC antiviral; antifungal; antirheumatic; antithyroid; antianemic. The  
CC sequences can be used for determining the presence of or predisposition  
CC to, or preventing or treating pathological conditions associated with an  
CC ORFX-associated disorder. The nucleic acids can be used to express ORFX  
CC proteins in gene therapy vectors. The proteins and nucleic acids may be  
CC used to treat cancers, proliferative disorders, neurodegenerative  
CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,  
CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester  
CC storage, systemic lupus erythematosus, severe combined immunodeficiency  
CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune  
CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and  
CC cartilage damage, nocturnal haemoglobinuria, antinflammatory disease; to  
CC enhance coagulation; to inhibit thrombosis; and as a contraceptive.  
XX  
SQ Sequence 2541 AA;

Query Match 2.4%; Score 224.5; DB 21; Length 2541;  
Best Local Similarity 19.4%; Pred. No. 5.6e-05;  
Matches 447; Conservative 299; Mismatches 843; Indels 713; Gaps 113;  
  
QY 11 KAAVHTAAHNPVG---HG-VALQ---QCSSSSSPQN-----AAASLAAEGKNGKMPR 56  
Db 428 kstviqqynrvqkvhegsvaipmrsgaspenfvqgmppaqqaitsgmhrghmnp 487  
QY 57 IQPSTFAADG-----ISAAHQKKSFLRCCLG-----TKFSPSAPQGPQTTHSK- 103  
Db 488 ltsaqagclgtinsmqavgaagatldtdtllpplqgdaasakwrknkmdeskeihsvq 547  
QY 104 -----GATRLDLARDDDGCTQHE-----AAAPDAARLTR-SGGVK-----RRNMDM--AGRP 148  
Db 548 daitagtasvnltagdpaetdytavgcavttissnltemsgvklalaledegsgsrp 607  
QY 149 MYKGGSGEDKVTQKRHLNFGQRMQLSKMAHPASANAGDRLOH-----SPPHGP 203  
Db 608 llaqaaglagavsellrsaqpasaeprqnlqaagvngqa-sgelliqigesdtdphfq 666  
QY 204 SHHEIKPEFVGSTKATTAHADRVETIAQDDDDSEFOQLHOORLARENPQPQKLGAT 263  
Db 667 almqi-akavasaalvllak--svaqrtedsglq---tqvlaaatqcalstsqvact 720  
QY 264 PISARFQKLTA-----VAESVLEGTDTTQSPKPSMLKSGAGVTPLAVT 310  
Db 721 kvva---ptisspvcqeglvagrivakavkcvsaqaatedgqllrgvgaat--avt 775  
QY 311 LDGKGLQALPDNPPALNTLLKOTGLKDTQHYLAHHASSDGSQHLLLDNKGHLFDIKSTAT 370  
Db 776 -----qalnell-----qhvkahatgagpagr-----yd-qatdt 804  
QY 371 SYSVLHN-----SHPGEIKGK---LAQAGT---GSVSVGKSGKISLGSTQSHNKTMLS 419  
Db 805 iltvtlenifsmgdagenvrqaarlaatsdlvnaikadae-----gesdlenstklis 858  
QY 420 QP---GEAHRSLLTGIWQHPAGAARPGQESIRLHDKDTHILHPELGWVSADKDTHSQLS 476  
Db 859 aakiladatakmv-----eaakgaa-----ahpd-----seeqqr 891  
QY 477 ROADGKLYALKDNRTLQNLSDNKSSEKLVDKTKSVSDQGOVALITDTFGRHKMSIMPS 536  
Db 892 eaacq---lrmatnaaq-----naikklivqrle-haakgaasatqtlaaqhaas-apk 943  
QY 537 LDASPESHISLSLH-FADAHQGLLHGKSELEAQSVVAISHGRVLWADSE-----GKLFFSA 589  
Db 944 asagqppllvqscakavaeqipllvqgvrgsqadpsaqlialiaasqsfllpgqkmvaa 1003  
QY 590 A---IPQGGDGNELKMKAMPQHALDEHPGHQHQISGFFHDDHGOQLNALVKN-----637  
Db 1004 akasvptlqg-----qasam-----qlsqcaknlgltalaelr 1035



PF 02-JUN-1999; 99WO-EP03848.

PR 03-JUN-1998; 98GB-0011962.

XX (JANC ) JANSEN PHARM NV.

XX Luyten WHML, De Raeymaeker MC, Geysen JUGH, Bogaert TAOB;

PI Maerten LJS, Verhasselt P, Van De Craen M;

XX WPI; 2000-116370/10.

DR N-PSDB; A07840.

XX Novel proteins and nucleic acids e.g. for treating neurodegeneration

XX Claim 3; Fig 1f; 146pp; English.

XX The invention provides vertebrate (human) protein homologue of a UNC-53  
 CC protein of *Caenorhabditis elegans*. The UNC-53 binds to microtubules or  
 CC their plus ends. The UNC-53 sequences are used to promote neural  
 CC regeneration, revascularization and wound healing; also for treating  
 CC neurodegenerative disease, acute traumatic injury, fibrotic disease and  
 CC autoimmune diseases (e.g. rheumatoid arthritis and sclerosis). The UNC-53  
 CC polynucleotides can be used for recombinant production of the proteins,  
 CC as a source of probes for detecting allelic variants and polymorphisms,  
 CC for sequencing genomic DNA and for detecting UNC-53 expression; and as  
 CC source of therapeutic antisense sequences. Cells that express the  
 CC protein are used to identify regulators of cell shape, growth, motility  
 CC and migration. They can also be used to identify proteins that are  
 CC involved in signal transduction pathways also involving UNC-53, and to  
 CC identify compounds that alter attachment of UNC-53 to microtubules. A  
 CC target gene coupled to a UNC-53 encoding sequence may be used to deliver  
 CC the target gene to a cellular microtubule or its plus ends. The present  
 CC sequence represents the amino acid sequence of the third human homologue  
 CC of UNC-53, designated hs-UNC-53/3.

XX Sequence 2385 AA;

Query Match 2.3%; Score 222; DB 21; Length 2385;

Best Local Similarity 18.5%; Pred. No. 7.3e-05;

Matches 405; Conservative 289; Mismatches 736; Indels 760; Gaps 106;

QY 22 VGHGVALQGGSSSSPONAAASLAEE-----GKNRGMPIRHQP-----STA 63

DB 203 vthasppeasqakqqdmqslsaryatqsnhsgiatsqkptrlpgrsrypaaqsssk 262

QY 64 ADGISAHQKKSFSL-----RCLGTKFERSAPQGPQTTHSKGATLRDLRLARDG 116

DB 263 vqgaslnrrsgsfnsidknkppnyangnekdsdkgppqssgv-----ng 307

QY 117 ETOHEAAA-----PDAARLTRSGVKRRNDDMAGRPMVKGGGEDKVPTQOKRH 166

DB 308 nvqppstaggppasaipsasakpws---ksmrvkhsatslmi----- 348

QY 167 QLNFGQMROTMKMAHPASANAGDRLOHSPPHIPGSHHEIKEPPVGSSTKATTAHADR 226

DB 349 -----tvkqgstasptps---drik--ppvseg---vktapsgaks----- 383

QY 227 VEIAQEDDDSEFQHLQORL--ARRENPPQPKLCV----- 261

DB 384 -----mlekfklnartalrpppppsgpgdgkddafesgemegfnsq 430

QY 262 ----ATPISARFQPKLT-----AVAESVLEGTD-----TTQSPLAQ--S 295

DB 431 nsggstusspksklappkagksnlnskkallqkekeeknrdnkvcetekpvkeekdq 490

QY 296 MLKGGGAGVPTPLVDKG-----KLQAPDN-----PPALNTLLKQTL-----GKD 337

DB 491 vtemapktksiaslipksgtktaakeslipsssggipkpgskvptvkqtispgstaske 550

QY 338 TOHYLAHHASSDGOHLLDNKHLFDIKSTATSVSVLHNSHPGEEKLA-----OAGTG 393

DB 551 sekfrtkgspqs-----lspitmekasascaplegreadqagpsgsc 597

QY 394 SVSDGKSGKISLGSST-----QSHNKTMLSOP--GEAHSLLTGIWQHAGAPARPQ 443

DB 598 tmtvaqssgq-stngavqlpqgqshsptatvapfiyrah-----senegtalps 648

QY 444 GESIRLHDDKTHILHPELGVWOSADKDTHSQLSRQADGKLYALKDNRTTQNLSD-NKSSE 502

DB 649 adsc-tsptkmdlsysk-----takqcleeisgeggpetrrmrtnvknialrqnl 697

QY 503 KLVDKIK----SYS-VDORGQVAILTDPGRHKMSIMPSLDASPESHISLSLHFAADAHQ 557

DB 698 etmslrlqtqishstlettftstvitvengr-----tipnltsrp---tpmcrwrlgacpr 750

QY 558 LLHGKSELEAQSVAISH-----GRLVADSEKGLFS-----AAIPKOGDGNELKMKAMPQ 607

DB 751 lqag----dapslgagyprrsgtsrftidpsrftmytptlrraavsrlgmssqidmsekas 806

QY 608 HALDEHFGHDHQSISGFFHDDHCOLNALYKNNFRQOHACPLGNDHQPFGNLTDLVIDN 667

DB 807 sldl--mssevdvvgymdsg-----dlilgs-----lrltd-----dinsgymtdg 844

QY 668 QLGLHHTN----PEPHEILDMCHGLSLALQEGKLLHYFDOLTKGWTGAESDCKOLKGLD 723

DB 845 glnltyrslnrtpdtatsrdi-----lqrg-vhdvtvdadssvssvglsdtdln 896

QY 724 AAYLLKGEVRLNINQSTSSI-----KHGTENVFSLPHVRNKPPEP 764

DB 897 is-----tdlnttssvssynitvpkrntqlrtidsekrsttdetwdspeelkkp 949

QY 765 ----GDA-----LOGLNKD-DKAAQAMAVIGVKN-----YLALTEKGDIRSFOIKPGTQ 807

DB 950 dfdshgdaggkwtvssglpepekagkasisvsqsgswrrgmagaqgapsrq-kagts 1008

QY 808 QLPERPAQT----LSREGISGELKDTHVDHKQNLALYALTHTHEGEVFPHPREAWGAESSWH 863

DB 1009 alktpgktdakasekg-kaplkqsslqrspsdagksdgckkppsglgrstatsf 1067

QY 864 KLALPOSESKLSDMSHEHKPIATFDGSOHLKAGWHAYAAPERGPLAVGTSGSQTV 923

DB 1068 k-----kpsgv---gssamitssg-----atitsgsatl 1093

QY 924 FNRLMQGVKGVIPGSGITVKLSAQTGTMGTAEGRK-----VSSKFSERI 968

DB 1094 -----gk-----ipksaigakksna-grktsldsgsqndvvlhvsskttlqy 1135

QY 969 RAYAFNPTMTPTPIKNAAYATQHGQWQREGIKPLIYEMOGALIKQLDAHNVHRNAPQD 1028

DB 1136 r-----slprpskst--sgiprgg-----hrstssi 1162

QY 1029 QSKLETLDLGEHGAEL-----LNDMKRFRDELEQSATRSVTVLGQHGV 1073

DB 1163 dsnvssksagattsklreptkigsgrspsvtvnqtdkekevavadsesvslsgspkssp 1222

QY 1074 KSNGEINSEFKPSPGK-----ALVQSFNVNRSQDLSKSLQOAVHATP--PSAESKL 1123

DB 1223 tsasacqaqglrppgskypdiastpfrlfgakagksasapntegvksvssvmpspstcl 1282

QY 1124 -----OSMLGHFVSAGVDMHQKEIPLGRQROPNDKT-----ALTKSLRLID 1166

DB 1283 arggslespspgtsgmsag---glsgssspifnk--psdltdvisishslaspasv 1337

QY 1167 TVTIGELHELADKAKLVSDHKPDADQIKLQQQFDTLREKRYESNPVKHYTDMGTNKA 1226

DB 1338 sftsgglvwaanmssssagsk-----dt---psyqsnstsh-----tsses 1375

QY 1227 LEANYDAVKAFINAFKKEHHG--VNLTRTVLESQSAELAKKLKNTLLSLDSGFSMSFS 1284

DB 1376 id-----lplshgslglttgt-hevqslmlrtgsvrstl-----sesmql 1417

QY 1285 RSYGGGVSTVFPTLSKKVPVPVPGAGITLDRAYNTLSFRTSGGLNYSFGDRGVSGNI 1344

DB 1418 rn-----tlpk-----glrytssrqanqeeqkew----- 1443

```
QY 1345 MVATGHDVMPYTKKTSAGNASDWLSAKHKISPDRLGIAAVSGTLQGTQNS---LKEK 1401
Db 1444 -----lrshstgqlgtgnqp-----lvsp-----samssaaagkyhfsnlvsptn 1485
QY 1402 LTEDLPQ-----FIHGLTHGTLTPAELLQKQIEH-----OMKQ 1435
Db 1486 lsgfnlpqsmrnsnpagdsfdydsqicgsatsleerpralshsgsfidsmeevh 1545
QY 1436 GSKLTFSDVTANLD-----LRAGINLNEDGSKPNGVTARYSAGLSASANL- 1481
Db 1546 gsslsivstsslsystaeeahseqihklrrelvasqe-----kvatitsqisanahlv 1599
QY 1482 AAGSRERSTTSGQFGSTTSASNNRPT-----FLNGVGAGANLTAALGVHAHSSTHE 1531
Db 1600 aafekslgmngirlqlmtaekeselielretiemlkaqnsaaq-aaigalngpdp 1658
QY 1532 GKPVGIFPAFTSTNVSAALDNRTSQSI-----SLELKRAEPTVNDISELTSTLG 1583
Db 1659 pkdlrirqhsesvs---sinsatshsigsgndadskkkknwnvnsrg-selrstrsfk 1714
QY 1584 KHKDSATTKMLAA---LKELDDA-----KPAEQLH-----1611
Db 1715 qafgkktkppshsdieeltdslpaepkphnagdcgsasmkpsqasasplvwppk 1774
QY 1612 -----ILQHFSAKDVVDERYEAVRNILKLVIRQO---AADSHSMELGSASHSTYNN- 1662
Db 1775 krqngpvlykhrsricecteaeeailqlkselrekelktldirlealsahhldqirea 1834
QY 1663 LSRINWDGIVELLHKHFD-----AALP-----ASSAKRLGEMWNN---1697
Db 1835 mnrmqne--feilkaendrlkaetgntakptrppssssssrqlsglsinnlnit 1892
QY 1698 DPALKDII-----KOLQSTPFSASVSMELKDGLE 1728
Db 1893 eaavsedillddagdatghkdgrsvkviivsisgkygrakqkqaylqgsi-----gvsg 1946
QY 1729 QTEKAILDGKVGGR--EEVGVLFDQDRNLNLRKVSVSQO-----SVSKSEGNFTPALLIGTSN 1782
Db 1947 ktkwdldgvlrllfkeyvfridrtstslgssdciasycigdlirshnlvpeell-----p 2002
QY 1783 SAAMSMERNIGTINFKYGODONTPRFETLE 1812
Db 2003 cgyivgdmniltvnlk-gveensldsfvfd 2031

RESULT 12
Y75096
ID Y75096 standard; Protein; 2048 AA.
XX AC Y75096;
XX DT 21-MAR-2000 (first entry)
XX DE Neisseria gonorrhoeae ORF 564 protein sequence SEQ ID NO:1666.
XX KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
XX KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
XX KW antibacterial; gene therapy.
XX OS Neisseria gonorrhoeae.
XX PN W09957280-A2.
XX PD 11-NOV-1999.
XX PF 30-APR-1999; 99WO-US09346.
XX PR 01-MAY-1998; 98US-0083758.
XX PR 31-JUL-1998; 98US-0094869.
XX PR 02-SEP-1998; 98US-0098994.
XX PR 02-SEP-1998; 98US-0099062.
XX PR 09-OCT-1998; 98US-0103749.
XX PR 09-OCT-1998; 98US-0103794.
```

```
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX (CHIR ) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI Petersen J, Pizzia M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;
XX WPI; 2000-062150/05.
DR N-PSDB; 253858.
XX Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics
XX Claim 2; Page 851; 1453pp; English.
XX 253015 to 254536, 254577 to 254615, and Y74253 to Y75941 represent
CC novel Neisseria meningitis and N. gonorrhoeae polynucleotides and
CC polypeptides. 254537 to 254576 and 254616 to 25473 represent PCR
CC primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.
XX SQ Sequence 2048 AA;
```

```
Query Match 2.3%; Score 221.5; DB 21; Length 2048;
Best Local Similarity 18.5%; Pred. No. 6.2e-05;
Matches 378; Conservative 255; Mismatches 763; Indels 645; Gaps 96;
QY 124 APDA-ARLTRSGGKVRNMDM-----AGRPWVGSGGDEKVPVQQRHQLNNFG 172
Db 277 aidangrlvsgtmaaanvqdmntaehkvnirsqafensgtavsgqgtqihsgsqiqntg 336
QY 173 QMRQTMLSKMAHPASAN-----AGDRL-----QHSPPHIPCSHHEIKE-----EPVG 214
Db 337 killsagtedlavsgslnnqngieatnqqlilhdgqgstvvidntngtigsgrdvaigaks 396
QY 215 STSKATTAAHADRVEITAEQDDDDSEFQQLHQORLARENPQPPLKGVATPISARFQPKLT 274
Db 397 lsnngtlaadnkldialqdd-----f 417
QY 275 AVAESVLEGDTTQSPKLPQSMKSGAGVTPPLAVTLDKGK-LOLAP---DNPPALNTLL 330
Db 418 yverktivagnels---lstrgslknsh-----tlqagkririkannldnavqngigs 466
QY 331 KOTLGKDTQHYLAHHASSDGSOHLL-----LDNKGHLFDIK 366
Db 467 ggttdigtqhltnrgllogqtkiqagqmnngitgriygdniataatrlndq-----den 522
QY 367 STATSYSLVHNSHPG-----EIKGKLAQAGTGSVSVD 398
Db 523 gtgaiaaarenlnlgieqlnnrenslisgndmavvgaldtdnqatgkqgrlhnagaile 582
QY 399 CKSKKISLG---SGTQSHNKTMLSQPGE-----AHSLLTGIWHPAG--AARP 442
Db 583 a-agkmrigveklhntnehlktqlvetgrerivdyeafrhellregt-qhelgwfvyynn 640
QY 443 QGESIRLHDDKTHILHPELGVWQSADKO---THSOLSROAGDKLYALKDNRTLQNLSDNK 499
Db 641 esdhlrtpdgvah-----enwhkydyekvtqetqvtgtapaklia-----680
QY 500 SSEKLVDKIKSYSDVOR-----COVAILTPTGRHKMSIMPSLDASPEHSISLSLHFADAH 555
```

Db 681 gsdliidskavfnsdsrliaggqllvqtckdglhneq----- 717  
QY 556 QGULLHGKSELAEQSVATISHGLVVDSEKGLFSAAIKPGQDGNELKWKAMPQHAIIDHFH 615  
Db 718 -----tfgekkyfsengklhnywrrarkghdetghreq-nytlpeesit 759  
QY 616 HDHQISGFFHDDHGLNALKVNNFRQHCACPLGNDHQFHPGNGNLTDLALVIDNQLGHLHT- 674  
Db 760 rdislgsfayeshsk--alsrhapsqgtelpqsrnd-----nirtak--sngislpyp 809  
QY 675 ---NPPHEIL-----DMGH-----LGSALQEGKLH-----Y 699  
Db 810 nsftipgsslyiinpankylvetdprfanyrqlwsgdymylgskldpnnlhkrlgdy 869  
QY 700 FDQ-----LT--KGWTGAESDCKQLKGLDGAAYLLKDGVEKRLNIN----- 739  
Db 870 yegrlneqiaetghrldgyndeeqfkalmndga-----taarsmnlsvglalsaeq 924  
QY 740 --QSTSSI-----KHGTENVFSLPHVRNKPDPGDALQGLNKDDKAQAMAVTCVN 786  
Db 925 aaqltsdiwlvqevklpdggtqtlv-mpqvyvrvkng-----gi--dgkallis--gsn 975  
QY 787 KYALATEKDIDRFQIKPGTQQLERPAQTLSREGISGEL-----KDIH-----VD 831  
Db 976 tqinvs--gslksgtiagnaliintdl--dnigrihaqsavtatqdinnigils 1031  
QY 832 HKONLYALTHEGEVFPHPREA--WQNGAESW--HKLALPQSESKLSLDMSEHKPIAT 888  
Db 1032 aeqlt--llnagninnqstaksnaggsstyldrmaglyitgkekgvlaaagqkdi 1089  
QY 889 FEDGSQHLKAGWHAYAPERGLAVTSGSQTVF-----NRLMOGVKGVIPGSGLTVK 944  
Db 1090 iaqqisnqsdqgqtrlagrdinldvtqkyqceihfidadnhtirgstnev--gssiqtk 1147  
QY 945 LSAQTGGMTCAEGRKYSSFE-----RIRAYAFNP--TMSTPRPIKNAAYATQHWQGR 997  
Db 1148 -----gdvtilsgnnlnakaevgsakgtlavaknditissgihagqvddaskh--tgr 1200  
QY 998 EGLKPLYEMOGALIKOLDANVHRHNAQPD--LQSKLETLDLGEHGAELLDNMRFRDELE 1056  
Db 1201 sg-----ggnklvltkdaqshetaqstfegkvvlqag-----nd----- 1237  
QY 1057 QSATRSVTVLGQHQGVKLSNGEINSEFKPSGKALVQSFVNNRSGDLKSLOQAVHATP 1116  
Db 1238 -----anilgsn--visdng-----trigaghnhrvrigttqtsqsetyhtq 1277  
QY 1117 PSAESKLQSLMGLHVSAGVDMSHQKGEIPLGRQDPNDKLTAKSLRLDVTIGELHEL 1176  
Db 1278 ksg-----lmsagifgtsgkntqngsgsnehtgstvgslkgdttivaskhye 1327  
QY 1177 ADKAKLVSDHKPDADQIKOLRQOFDTLREKRYESNPVKHYTDMGFT----- 1222  
Db 1328 qtgsnvss---pegnnlistqsmidgaacnqlnskttqyeqkltvgiqfaryrfgtts 1384  
QY 1223 -----HNKALEANYDAVKAFINAFKKEHGVNLTTRVLESQGSAAELAK-----KLNKTL 1272  
Db 1385 dcrstqsakqvqgskndrvnamaaa-----nagwqayvtqkgagqnlnganttnakqvs 1436  
QY 1273 LSLDSGESMS-----FSRSYGGGVSTVFVPTLSKKVPVPVI-----PCAGI 1313  
Db 1437 isitygeqnrqtqvtqanqaqasqiaqagkttlycrrcgqsninitsgsvsgragtgl 1496  
QY 1314 TLDRAYNLSFSTMS-----GGLNYSFCRDGGVSGNIMVATGHDVMPYMTGKK 1360  
Db 1497 iadkqihlqsaesqntersqnksagwnagaavsfgggwslg---vaaggnv-----gkg 1548  
QY 1361 TSAGNASDWLSAKHKISPD-----LRTGAASVG----- 1388  
Db 1549 yyyg---dsvthrshidgksqtlisggdttikgaqvrkgvqvnaknlsiqsvqdre 1605  
QY 1389 TLQGTFLONS-----LKFKLTDELPGFIHGLTHTLPABELQKIE--HOMKQGSKLT 1440  
Db 1606 tyqskqnaqaqvtvgvfyfsagdyssqskira-dhasvtceqsgiyagdygqikvqn--- 1661

QY 1441 FSVDTSANLDRAGIINLEDGSKPNGVTARVSAGLSASANLAAGSRERSTTSQGFSTTS 1500  
Db 1662 -----hcgkkgiitssqsakdkgn-rfstqtlagsdiqnyseqegksfg-lgasva 1712  
QY 1501 AS-----NNRP-----TFLNG-----VGAGANLTAALGVVAHSTHEGKVPVIFPA 1540  
Db 1713 vsqktlqggaaknpqdkhltiadkngasssvgygsdsqssitksgintpknqi--- 1769  
QY 1541 FTSTNVSAALALDNRTSQSTISLELKRAEPTVSTNDISELTSTLGHFKFDSATTKMLAALKE 1600  
Db 1770 ---tdaaagirltgkiaaqtka---dtntvtdtaerh-----sgslkn 1808  
QY 1601 LDDAKPAEOLHIQ-----QHFSADVVDERYEAVRNLLKLVIRQQAADSHSME--LGSAS 1655  
Db 1809 ifdkdrvsqeldlqtrvtsqfskn--vqqtnteinqhlhdkkadkeaaetaaaalaangd 1866  
QY 1656 HSTTYNNLSRINDGIVELLHKHFDALPASSAKRKLGEEMNN-----DPALKDI--IKQ 1707  
Db 1867 metakrk-----aheaqdaaakadnwqg-gkvlmnlmlasglaeptqsgaglaa 1913  
QY 1708 LQSTPFSSASVSMELKDLREQTEKAILDGKVGREEVGLFQDRNNLRVKSVS--SQSVS 1766  
Db 1914 ataspdvsyiaighfkf-----lagngangkitasqetahviahavlgava 1960  
QY 1767 KSEGFNTPALLIGTSSAAMSERNITGTINFKYQDQDQNTPRFTLEGGIAQANPQVASAL 1826  
Db 1961 aaxgnnapagalaggseaaa--piig--kwylykgdg-----gslnaeeketvsai 2008  
QY 1827 T 1827  
Db 2009 t 2009  
RESULT 13  
R05041  
ID R05041 standard; protein; 3647 AA.  
XX R05041;  
XX 02-OCT-1990 (first entry)  
DT Filamentous haemagglutinin A.  
DE Filamentous haemagglutinin; fhaB; whooping cough; vaccine.  
KW Filamentous haemagglutinin; fhaB; whooping cough; vaccine.  
OS Bordetella pertussis.  
XX Key Location/Qualifiers  
FH Region 1..1069  
FT /label=N-terminal  
FT /note="highly basic; 98 kD"  
FT Region 1074..3647  
FT /label=C-terminal  
FT /note="acidic and hydrophilic; 140 kD"  
FT Cleavage-site 1069..1073  
FT /label=trypsin-like proteolytic cleavage site  
FT /note="confirmed by SDS-PAGE peptide analysis."  
FT Binding-site 1097..1099  
FT /label=cell recognition site  
FT Binding-site 2651..2653  
FT /note="surface exposed, hydrophilic, antigenic"  
FT /label=cell recognition site  
FT Region 2000..2700  
FT /label=high concn. of charged residues  
FT Region 1800..2000  
FT /label=hydrophobic region  
FT Region 2400..2500  
FT /label=hydrophobic region  
FT Domain 44..69  
FT /label=transmembrane helix  
FT Region 52..69  
FT /label=transmembrane segment of helix







Db 1801 entakls--gevrkqvqdvqggehrgwsgigvnywvrlragngkkaagtiaapwyggdita 1858  
QY 443 -----QGESIRLH-----DDKIHILHPELGVWQSDAKDTHSLSROADCK----- 482  
Db 1859 eqslievqkdlvlnagarkdehrllne--gviqagg--hghiggdvdnrvvrtvsam 1913  
QY 483 -----LYALKDNRLQNLSDN--KSEKLVDRKTSYSDORGVQVAILTDPGRHKM 531  
Db 1914 eyfktplvpvltalndnraglspatwfnfstyeld-----yldqnryeylwglyptyctew 1969  
QY 532 SI-----MPSLDASPEHISLSLHPADAHQ-----LLHG-----KS 563  
Db 1970 svntlknlidvgakpaptappmpkape--ldirghtlesaeqrkifgeykkkgleyeka 2027  
QY 564 ELEQASV-AISHGRLVVAQSEKFLSAI-----PKQDGNELMKAMPOHALDEHFGH 616  
Db 2028 knavqaveaygeatrivrhdqglgrygkalgmdaetkevq-----liqefaa 2075  
QY 617 DHQISGFFHDDHQLNA---LYKNFRQO--HACPLGNDHOPHPCWNLTDLALVIDNOLG-- 670  
Db 2076 dirtvyakadqatidaetkvaqryksqidavrli---qalqpy-rvllakaisaalgaad 2131  
QY 671 ----LHHTN-----PEPHEILDMGHLSLALOECKLHYFDQLTKGW 707  
Db 2132 wralghsqimqrwkdfkagrgaeiafykeqvtlaag--agltlsngaih-----nge 2183  
QY 708 TGAESCKO--LKKG-----LDGAAYLLKD-GEVKRLINQSTSSIKHGTEVPSLPHV 758  
Db 2184 naeqnrgrepglkighatsvsgsfalrdvglekrldiddalaa-----vlnvphi 2236  
QY 759 -----RNKPE-----PGDALOGLNKKDKAKAMAVI 783  
Db 2237 frigaqaqsladgaagpallargaqapetdgmvdarglsadalasiasdaaglevs 2296  
QY 784 GVNKYLALTEKGDIRSFQIKPGTQOLERP---AOTLSREGISGELKD---IHVDHKONLY 837  
Db 2297 g-rinaqvadag-----lagpsavaapavgaadvgepvtgdvqdvppvavaglegpva 2348  
QY 838 ALTHEGEVHPQREAMQNGAESWHLALPOSESLSKLSLDSMHEHKPIATFEDGSQHOL 897  
Db 2349 tvrvappavalpr-----plfetrifikdqsky-----gsryff 2383  
QY 898 KAGGWHAYAPERGPLAVGTSGSOTVF-----NRLMQVKQKV-IPGSLTVKLKSAQGG 951  
Db 2384 eqigy-----kpdraarvagdyfttlvreqvrralgyesrlpvrvalvaklmdsadt 2439  
QY 952 MTGAERKV-----SSKFSERIRAYAFNPTMTTPRIKNAAYATOHGWQREGKPLY 1004  
Db 2440 vkgalgvkvaptaqklkqadrfvvyvtdvqkvlaprlyte---atrqgltddy 2496  
QY 1005 EMOCALIK-----QLDAHVRHNAPOPDLQSKLETLDLGHGAELNDMKRFRDELEQ 1057  
Db 2497 agggalliasgddvtvntdghdv---ssvngllgrsvkvdaqg-kvrvvadskgagggie- 2552  
QY 1058 SATRSVTVLGOHQGVLSKNGEINSEFPKSPKALVOSFNVRNSGODLSKSLQOAVHATPP 1117  
Db 2553 -addevdvsgrdigi--egglr-----gkdvrlikadvkvatsm 2589  
QY 1118 SAESKLSQMLGHFVSAGVD--MSHOKGEIPLGRORPNDKALT--KSRLILDTVTIGELH 1174  
Db 2590 ryodk-----griaardgaladagggqllhieakrlretagatlkggkvklvdvdklgyvy 2644  
QY 1175 ELADKAKLVSDHKPDADQIKOLROQFD7LREKRYESNPVKHYTDMGFTH---NKALEANY 1231  
Db 2645 eagss-----yen---ksstplgslfailststetnq 2673  
QY 1232 DAYKAFINAPKKEHHGVNLTTRTV-----LESQSAELAKKLNTLLSLDGSMSFSR 1285  
Db 2674 sa-----hanhygrtrieagtleqknqleieggsvda---ahtdlsvardarfkaaa 2722  
QY 1286 SYGGGVSTVFPTLSKKVPVPIPGAGITLDRAYNLFSRSTS--GGLNVSFGRDGVGSGNI 1344  
Db 2723 dfahaehkdvrvqlslgakv---gag-----gyeagfsalgsesgleahagr--gmtaga 2771

QY 1345 MVATGHDVMPYMT-----GKTSAGNASDWLSAKHKISPDL---RIGAAV 1386  
Db 2772 evkvgyrasheqsseteksyrynannlnfggsvaeagnvldigga-----dinrnyggaa 2825  
QY 1387 SGTLOCTQLNSLKFTEDELPGFIHGLT-----HGTLTFAELLQKGI 1429  
Db 2826 kgn-agt-eealrmrakvestkyvseqtsgsgswsvastasarssltaatrldgsv 2883  
QY 1430 EHQMKQGSKL-----TPSVDTSANLDRAGINLNEGSGPNGVTA----- 1469  
Db 2884 agnvedgreirgelmaaqvaaeatqlvctadtaa--valsagisadfdshsrstsqntqyl 2942  
QY 1470 -----RVSAGLSASANLAAGSR--ERSTTSQCGFSTTSASN 1503  
Db 2943 ggnlsieategdatlvgakfgggdqvsikaaksvnlmaaeestfesyshnhasadanl 3002  
QY 1504 NRPTFLNGVGAGANLTAALGVVAHSTHE-GRPVGIFPAPTSTNVSA--LALDNRTSQSIS 1561  
Db 3003 ganavggavglg--ltagmgtshqitnetgk-----tyagtsvdaanvsid--agkdln 3052  
QY 1562 LELKRAEPVTSNDISELTSTLGHKFKDSATTKMLAALKELDDAKPAEQLHILQOHFSKAD 1621  
Db 3053 lsgsrivr-----gkhvvldevdinatskq----- 3077  
QY 1622 VVGDER-YE-----AVRNKKLVIRQQAADSHSMELGSA--SHSTTYNNLSRI 1666  
Db 3078 ---derynssgggwdasagvaigq-rtlv-----apvgsagfnftehdn--srl 3122  
QY 1667 NNGIVBELL-----HKHFAALPASSAKRKGEMNNDPALKDIKLOLQSTPFSSASVS 1719  
Db 3123 tndgaagvvasdgitghvkgdanitgatdiadlsgkgn-----lk-----vdgavna 3168  
QY 1720 MELKDGREOTEKAILDGKVGREEVGLFQDRNRLRVKSVSVQSVKSEGFNTPALLLG 1779  
Db 3169 qnlkd-yrdk-----dggsgglvng-----issttlaptvgvafg-----rvag 3206  
QY 1780 TSNSAASMERNIGTINFKYQDQNTPRRFTLEGGIAQANPQVASALTDLKK 1831  
Db 3207 edyqaedratidvg-----qtkdparlvqggvgvgtlngdaaqtavvqr 3250

RESULT 15  
W04505  
ID W04505 standard; Protein; 1992 AA.  
XX  
AC W04505;  
XX  
DT 25-JAN-1997 (first entry)  
XX  
DE Moraxella 200 kDa outer membrane protein.  
XX  
KW Outer membrane protein; OMP; immunogen; vaccine; otitis media;  
KW diagnosis.  
XX  
OS Mycobacterium catarrhalis strain 4223.  
XX  
FN W09634960-A1.  
XX  
PD 07-NOV-1996.  
XX  
PF 29-APR-1996; 96WO-CA00264.  
XX  
PR 26-MAR-1996; 96US-0621944.  
PR 01-MAY-1995; 95US-0431718.  
PR 07-JUN-1995; 95US-0478370.  
XX  
PA (CONN-) CONNAUGHT LAB LTD.  
XX  
PI Chong P, Harkness RE, Klein MH, Loosmore SM, Sasaki K;  
XX  
DR WPI; 1996-506162/50.  
DR N-PSDB; T38740.

xx Moraxella outer membrane protein - useful as immunogen in protective  
 PT vaccine and for diagnosis

xx Claim 14: Fig 6; 109pp; English.

cc An approx. 200 kDa outer membrane protein (W04505) can be  
 cc isolated from Moraxella catarrhalis otitis media strain 4223  
 cc by electroelution, or expressed from a gene (see also F38740)  
 cc obt'd. from a strain 4223 genomic library. Natural or recombinant  
 cc outer membrane protein is useful as an immunogen to protect  
 cc against infection by Moraxella, esp. M. catarrhalis. It can  
 cc also be used to detect antibodies, esp. for differential diagnosis  
 cc between bacteria that cause similar symptoms, and also useful as  
 cc a carrier for other antigens and used to raise antitumour  
 cc antibodies for conjugation to therapeutic agents.

xx Sequence 1992 AA:

Query Match 2.28; Score 204.5; DB 17; Length 1992;  
 Best Local Similarity 19.08; Pred. No. 0.00071;  
 Matches 316; Conservative 221; Mismatches 606; Indels 531; Gaps 79;

QY 469 KDT-HSOLSRQ-----ADG-KLYALKDNRTLN-----LSDNKSSEKLVDRKIS 510  
 DB 15 kdtkhaigeqnprsgtakadgralaigenanaagqgaiaigsnt-----vng 67  
 QY 511 YSDORQVAILTDPGRHKMSIPSLDASPESHISL---SLHFDADAGLHUKSELEA 567  
 DB 68 ssldkig-----datgesaigdvkaskdasiaigsddlhldqgnphkpgktlin 122  
 QY 568 QSAVISHGRLVADSECKLFSAAPKOGDGNELKMKAMPQHALDEHFGHDHIGSFFHDD 627  
 DB 123 dli---nghavkvei-----rsskndvkyrrtas-----ghastavgamsya 163  
 QY 628 HGQNALVKNFNROOHACPLGNHDHGFHPGNWLTDALVIDNQLGHLHTPEPEHILDMG-- 685  
 DB 164 qghfs---naf-----gtrataksayslavglaat-aeggstiaigsd 202  
 QY 686 ----HLGSLALOG---KLHYFDQLTKGWTGAESCKOL-----KGLDGAAYLLKQGE 732  
 DB 203 atsslgalaigagtraqlgsgialgsgsvvtqsdnnsrpaytpntqaldpdkfgatnntk 262  
 QY 733 VKRLINQSTSSIKHGTENFSLPHVRNKPPEGDALGNLKDCK---AQAMAVIGVKNYL 789  
 DB 263 agplsi--gnsirrkiniaga-----gvnktadvnvaqleavvkwaker 305  
 QY 790 ALTEKGDIRSFQIKPGTQO-----LERPAQTLSREGISGELKD--IHVDHKQNLALY 840  
 DB 306 rltfggdndstdvkigldntltikggaetnaltldnnigvkvkeadnglklaktlnnlt 365  
 QY 841 HEGEVFHQPREAWONGAESSEWHKIAL-----POSESK-----LKSLDMSHE 882  
 DB 366 evnttlnattvkvgssttaellsdslftqntgsgtsktvgyvngvkvfnnnaet 425  
 QY 883 HKPIAT-----FEDG-----SOHOLKAG-----GWHYAAAPERGPLA 914  
 DB 426 taaigttrirtdkigardgvdekqapyldkqlkvgsvaitidngida-gnkkislnla 484  
 QY 915 VGTSGSTVFNRLMGQVKGKVIPOSGLTV-----KLSAQOTGGM-----GAE 956  
 DB 485 kgsandavtieqlkaakptinagagisvptseisvdaksngvntaptynigvkttelnsd 544  
 QY 957 GRKVSSKFSERIRAYANPMTSPRPI----- 983  
 DB 545 g--tsdkfs--vksgstnnsivtaehiaslyinevnrtdadsalsqsfvkeedddanaity 600  
 QY 984 -----KNAAYATQHWGQREGLKPLXYEQGALIKQL-----DAHNVRHNA 1023  
 DB 601 akdtknagavsiiklkgngkltvatkdkdgtvtfsgdsqsltigkstlndnlgltvkdn 660  
 QY 1024 POPDLOS---KLETLDLGEGHGAELLNDMKRPRDELQESATRSVTVLGGHQGVLSNGEIN 1080

DB 661 eqiqvgangakftnvnsgnpgtgiantaritrdkifags-----dgavdnkpyl 711  
 QY 1081 SEFKPSGKALVOSFNVRNSQDLSKSLQQAQVHATPPSAESKLSQMLGHFVSAGVDMSHO 1140  
 DB 712 dqdklqvgnvkiitntginaggkait-----glsptlpsi-----adqqs 751  
 QY 1141 KGEIPLGRQRPNDKTKALTAKSRLLIDT-----VTIGEL 1173  
 DB 752 --nieigtigdkdknaasindilntgfnlknmnpidfsvtydivdfangnattatvt 809  
 QY 1174 HELADK-AKLVSD-----HKPDADQIKOLROQFDLREKRYESNPVKHYTDMGFTHN 1224  
 DB 810 hdtanktskvyvdyvnddttihltgtdnkkigvkttklntsangn-----tatfnvn 864  
 QY 1225 KALE---ANYDAVKAFINAFKKEHGVNLTTRTVLESQGSABELAKKLNTLLSLDSGESM 1281  
 DB 865 ssdedalvnaadiaenintlakeihktgtadtaltq-----ftvkkvdenndannait 920  
 QY 1282 SFSRSYGGGYSTVFPVPTLSKKVPVPVPCAGITLDRAYNLNLSFS-RTSGGLNV--SFRDGG 1338  
 DB 921 vdgknannqvnltl---tlkge-----nglniktdkngvtfginttsglkagkstinlg 971  
 QY 1339 GVS-----GNIMVATGHDVMPYMTGKKTSAQNADSLSAKHKISPDLRIGAAVSGTLQGT 1393  
 DB 972 glsiknptgseqiqvgadgvkfa-----kvnnngvvvgagidgt--- 1010  
 QY 1394 LQNSLFKLTDELPGFIHGLTHGTLTPAE--LLOKGEIHOHQKQSKUTFSVDTSANL 1451  
 DB 1011 -----ritrdei-ftgt--tngsldkspkhlsgkdnin--aggkkit-----ni 1048  
 QY 1452 RAG-INLINE---DGSKPNGVTARVSAGLSASANLAAGSRERSTSGOFGSTTSANNRP 1506  
 DB 1049 qsgeiaqshdvtvggiydlkteleknissaktagnslhefsvadeqgnnftvsnpy 1108  
 QY 1507 TF-----LNGVGAGANLTAALGVAHSSTHEGKPGVIGIPPAFTSTNVSAALAD 1553  
 DB 1109 sydtsktsdvittagengittkvn---kgvrvrigdtkgtltpkltvgnngkgivid 1164  
 QY 1554 NRTSQSISLEKRAEPVTSNDISELTSTLGHKHPKDSA---TTKMLAALKELDDAKPABOL 1610  
 DB 1165 sqngq-----ntifglsntlanvtnndkgsrvrttegniikdedktraasiv 1210  
 QY 1611 HILOQHFS-----AKDVVGDERVEAYRNKLKLVIRQQAADSHSMELGSASHSTYNNLS 1664  
 DB 1211 dvlsagfnlqngaeavdvf--stydtv-----nfadgn-----attakvlyddts 1253  
 QY 1665 RIN-----NDGIVELLHKHF-----DAALPAS----- 1686  
 DB 1254 ktskvydyvnnvddttievdkkkgvktltstgtgankfalsnqatgdalvkaasdivah 1313  
 QY 1687 -----SAKRLGEMNNN-----DPALKDIIKOLQSTPPSSASVSMELKGLRQTEKA 1733  
 DB 1314 lntlsdgiqtakgasqannsgayvddgngkviydstdnkyygak-----ndgtvdktkv 1368  
 QY 1734 ILDGKVGREVE--GVLFODRNRLNRVKSVSVSOSV---SKSEGFNTP-ALLLGTNSAAMS 1787  
 DB 1369 akdklvagactpdgttlaq---mnvksinkevndankkqginednafvkglekaasdn 1424  
 QY 1788 MERN-----IGTIN-----FKYGQDQNTPRR-----FTLEGGIAQANPOVASALTD 1828  
 DB 1425 ktknaavtvgdlnavaqtpltfagdtgtaktakkgititkggqtdn-----kltd 1475

Search completed: June 5, 2001, 18:16:36  
 Job time: 202 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2001, 18:23:10 ; Search time 89.87 Seconds  
(without alignments)  
181.283 Million cell updates/sec

Title: US-09-596-784-4

Perfect score: 715

Sequence: 1 MTSSQORVERFLQYFAGCK.....IEHAAEVREYIAOLDESSAA 139

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_15:  
1: sp-archaea:  
2: sp-bacteria:  
3: sp-fungi:  
4: sp-human:  
5: sp-invertebrate:  
6: sp-mammal:  
7: sp-mhc:  
8: sp-organelle:  
9: sp-phage:  
10: sp-plant:  
11: sp-rodent:  
12: sp-unclassified:  
13: sp-vertebrate:  
14: sp-virus:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	715	100.0	139	2	O54621 erwinia amy
2	455	63.6	124	2	O9KH43 erwinia her
3	273.5	38.3	129	2	O66102 pseudomonas
4	273.5	38.3	131	2	O9JP37 pseudomonas
5	80.5	11.3	368	10	O40879 Q40879 pellargonium
6	80.5	11.3	766	4	O9UKF4 homo sapien
7	80.5	11.3	787	4	O9UKF3 homo sapien
8	80.5	11.3	820	4	O9UKF5 homo sapien
9	80.5	11.3	820	4	O9UHP1 homo sapien
10	78.5	11.0	830	2	O9PE32 homo sapien
11	77.5	10.8	775	2	O9X7B0 xylella fas
12	77	10.8	1166	4	O9PE33 mycobacteri
13	75.5	10.6	402	5	O45260 caenorhabdi
14	75	10.5	2053	5	O9XYM2 caenorhabdi
15	75	10.5	2053	5	O9W5N4 drosophila
16	73	10.2	578	11	O62693 rattus norv
17	72.5	10.1	2183	14	O98588 subacute sc
18	72	10.1	377	10	O9LH84 arabidopsis
19	71.5	10.0	1591	4	O92601 homo sapien

20	71.5	10.0	2484	5	O9U347 caenorhabdi
21	71.5	10.0	2607	5	O23187 caenorhabdi
22	71	9.9	546	11	O9J124 rattus norv
23	70.5	9.9	1048	5	O9NA71 caenorhabdi
24	70.5	9.9	2183	14	O98334 measles vir
25	70.5	9.9	2183	14	O98589 measles vir
26	70.5	9.9	2183	14	O9WPy9 mus musculu
27	70	9.8	594	11	O9JID5 watermelon
28	70	9.8	682	14	O56135 coccidioid
29	70	9.8	1860	3	O9Y878 caenorhabdi
30	69.5	9.7	291	5	O9XTR0 cephalospor
31	69.5	9.7	359	8	O37749 synchocyst
32	69.5	9.7	393	2	O55689 measles vir
33	69.5	9.7	2183	14	O83626 measles vir
34	69.5	9.7	2183	14	O9YZ23 measles vir
35	69.5	9.7	2183	14	O9YZY9 measles vir
36	69.5	9.7	2183	14	O9WP21 measles vir
37	69.5	9.7	2183	14	O9WPY7 measles vir
38	69.5	9.7	2183	14	O9WPY8 measles vir
39	69.5	9.7	2183	14	O9WMB3 measles vir
40	69.5	9.7	2183	14	O9WPY6 measles vir
41	69.5	9.7	2183	14	O9WP20 measles vir
42	69.5	9.7	2183	14	O9W984 measles vir
43	69.5	9.7	2183	14	O9IFK4 measles vir
44	69.5	9.7	2183	14	O9IC35 measles vir
45	69.5	9.7	3680	6	O97592 canis famil

## ALIGNMENTS

## RESULT 1

O54621 ID O54621 PRELIMINARY: PRT; 139 AA.  
AC O54621;  
DT 01-JUN-1998 (TREMREL. 06, Created)  
DT 01-JUN-1998 (TREMREL. 06, Last sequence update)  
DT 01-NOV-1998 (TREMREL. 08, Last annotation update)  
DE DSPF.  
GN DSPF OR DSPF.  
OS Erwinia amylovora.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Erwinia.  
OX NCBI\_TaxID=552;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=EA321;  
RX MEDLINE=98115919; PubMed=9448330;  
RA Bogdanov A.J., Kim J.F., Wei Z., Kolchinsky P., Charkowski A.O.,  
RA Conlin A.K., Collmer A., Beer S.V.;  
RT "Homology and functional similarity of an hrp-linked pathogenicity  
locus, dspEF, of Erwinia amylovora and the avirulence locus avrE of  
Pseudomonas syringae pathovar tomato.";  
RT Proc. Natl. Acad. Sci. U.S.A. 95:1325-1330(1998).  
RL [2]  
RN SEQUENCE FROM N.A.  
RP STRAIN=EA321;  
RC Bogdanov A.J., Kim J.F., Wei Z., Kolchinsky P., Beer S.V.;  
RA Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.  
RL [3]  
RN SEQUENCE FROM N.A.  
RP STRAIN=CfBP1430;  
RC MEDLINE=98086111; PubMed=9426142;  
RA Caudrault S., Malandrin L., Paulin J.P., Barny M.A.;  
RT "DspA, an essential pathogenicity factor of Erwinia amylovora showing  
homology with AvrE of Pseudomonas syringae, is secreted via the Hrp  
secretion pathway in a DspB-dependent way.";  
RL Mol. Microbiol. 26:1057-1069(1997).  
DR EMBL; U97504; AAC04851.1; -  
DR EMBL; Y13831; CAAT4157.1; -  
SQ SEQUENCE 139 AA; 15723 MW; 60F23F88B1BFAE93 CRC64;

RT locus, dspef, of Erwinia amylovora and the avirulence locus avrE of  
 RT Pseudomonas syringae pathovar tomato.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 95:1325-1330(1998).  
 DR EMBL: U97505: AAC06135.1; -  
 SQ SEQUENCE 129 AA; 14339 MW; 19ABBD84968ECE9 CRC64;

Query Match 38.3%; Score 273.5; DB 2; Length 129;  
 Best Local Similarity 43.1%; Pred. No. 1.7e-21;  
 Matches 56; Conservative 27; Mismatches 44; Indels 3; Gaps 3;

Qy 1 MTSSQORVERFLOYFSAGCKTPHILKDGVCALYNEDEEAALVLEVPQHSLSLLHCRRIE 60  
 Db 1 MTSSQORVERFLOYFSAGCKTPHILKDGVCALYNEDEEAALVLEVPQHSLSLLHCRRIE 60  
 Qy 61 ADPQTSITLYSMLQLNFEMAAARGCWALDELHNVRCLFQOQSLHDLDEASFSDIVSGFI 120  
 Db 61 ADPQTSITLYSMLQLNFEMAAARGCWALDELHNVRCLFQOQSLHDLDEASFSDIVSGFI 120  
 Qy 61 C-PERAPDLL-RLLSINFDVARLHGCWFAVDQ-GDVRCLCAQRELASLDEPAFCDVTRGFI 117  
 Db 121 EHADEVREYI 130  
 Db 118 SQAREARAF 127

RESULT 4  
 Q9JP37 PRELIMINARY; PRT; 131 AA.  
 AC Q9JP37;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)  
 DE HYPOTHETICAL 14.6 KDA PROTEIN.  
 OS Pseudomonas syringae (pv. tomato).  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 OX NCBI\_TaxID=323;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=DC3000;  
 RX MEDLINE=98422476; PubMed=9748456;  
 RA Charkowski A.O., Alfano J.R., Preston G., Yuan J., He S.Y.,  
 RA Collmer A.;  
 RT "The Pseudomonas syringae pv. tomato HrpW protein has domains similar  
 RT to harpins and pectate lyases and can elicit the plant hypersensitive  
 RT response and bind to pectate."  
 RL J. Bacteriol. 180:5211-5217(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=DC3000;  
 RX MEDLINE=20243785; PubMed=10781092;  
 RA Alfano J.R., Charkowski A.O., Deng W.L., Badel J.L.,  
 RA Petnicki-Ocwieja T., van Dijk K., Collmer A.;  
 RT "The Pseudomonas syringae Hrp pathogenicity island has a tripartite  
 RT mosaic structure composed of a cluster of type III secretion genes  
 RT bounded by exchangeable effector and conserved effector loci that  
 RT contribute to parasitic fitness and pathogenicity in plants."  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:4856-4861(2000).  
 RL EMBL: AF232006; AAF71500.1; -  
 DR Hypothetical protein.  
 KW SEQUENCE 131 AA; 14571 MW; F2A34660EF98EC98 CRC64;

Query Match 38.3%; Score 273.5; DB 2; Length 131;  
 Best Local Similarity 43.1%; Pred. No. 1.8e-21;  
 Matches 56; Conservative 27; Mismatches 44; Indels 3; Gaps 3;

Qy 1 MTSSQORVERFLOYFSAGCKTPHILKDGVCALYNEDEEAALVLEVPQHSLSLLHCRRIE 60  
 Db 3 MKTSQDPDFARFINSGLAQGLTSLTQNGVLCALYDGNNEAIIELPEHSEWVIFHCRIGR 62  
 Qy 61 ADPQTSITLYSMLQLNFEMAAARGCWALDELHNVRCLFQOQSLHDLDEASFSDIVSGFI 120  
 Db 63 C-PERAPDLL-RLLSINFDVARLHGCWFAVDQ-GDVRCLCAQRELASLDEPAFCDVTRGFI 119

Query Match 100.0%; Score 715; DB 2; Length 139;  
 Best Local Similarity 100.0%; Pred. No. 4.3e-68;  
 Matches 139; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MTSSQORVERFLOYFSAGCKTPHILKDGVCALYNEDEEAALVLEVPQHSLSLLHCRRIE 60  
 Db 1 MTSSQORVERFLOYFSAGCKTPHILKDGVCALYNEDEEAALVLEVPQHSLSLLHCRRIE 60  
 Qy 61 ADPQTSITLYSMLQLNFEMAAARGCWALDELHNVRCLFQOQSLHDLDEASFSDIVSGFI 120  
 Db 61 ADPQTSITLYSMLQLNFEMAAARGCWALDELHNVRCLFQOQSLHDLDEASFSDIVSGFI 120  
 Qy 121 EHADEVREYIAQDESSAA 139  
 Db 121 EHADEVREYIAQDESSAA 139

RESULT 2  
 Q9KH43 PRELIMINARY; PRT; 124 AA.  
 AC Q9KH43;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)  
 DE DSPF.  
 GN DSPF.  
 OS Erwinia herbicola pv. gypsophilae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Erwinia.  
 OX NCBI\_TaxID=48984;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Mor H., Manulis S., Zuck M., Nizan R., Coplin D.L., Barash I.;  
 RT "Genetic Organization of the hrp Gene Cluster and dspef Operon in  
 RT Erwinia herbicola pv. gypsophilae."  
 RL Submitted (May-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF271717; AAF76344.1; -  
 SQ SEQUENCE 124 AA; 13873 MW; 9F0266FCB78CCEB0 CRC64;

Query Match 63.6%; Score 455; DB 2; Length 124;  
 Best Local Similarity 67.5%; Pred. No. 1.1e-40;  
 Matches 81; Conservative 18; Mismatches 21; Indels 0; Gaps 0;

Qy 1 MTSSQORVERFLOYFSAGCKTPHILKDGVCALYNEDEEAALVLEVPQHSLSLLHCRRIE 60  
 Db 1 MTPSQORVDAFLOHFAAACKTPIHLKEGVCVPMITDQGEAAALVLEVPQHSLSLLHCELAS 60  
 Qy 61 ADPQTSITLYSMLQLNFEMAAARGCWALDELHNVRCLFQOQSLHDLDEASFSDIVSGFI 120  
 Db 61 TAPHSALNFYALLQLNFEMAAARGCWALDELHNVRCLFQOQSLHDLDEASFSDIVSGFI 120

RESULT 3  
 Q66102 PRELIMINARY; PRT; 129 AA.  
 AC Q66102;  
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)  
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
 DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)  
 DE AVIRULENCE PROTEIN.  
 GN AVRF.  
 OS Pseudomonas syringae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 OX NCBI\_TaxID=317;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=DC3000;  
 RX MEDLINE=98115919; PubMed=9448330;  
 RA Bogdanove A.J., Kim J.F., Wei Z., Kolchinsky P., Charkowski A.O.,  
 RA Conlin A.K., Collmer A., Beer S.V.;  
 RT "Homology and functional similarity of an hrp-linked pathogenicity





```
QY 119 FIEHAAEVREYIAQLDESS 137
ID Q9UKF5 PRELIMINARY; PRT; 820 AA.
AC Q9UKF5;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE METALLAPROTEINASE-DISINTEGRIN.
GN ADAM29.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Cerretti D.P., DuBoise R.F., Black R.A., Nelson N.;
RT "Isolation of Two Novel Metalloproteinase-Disintegrin (ADAM) cDNAs
RT That Show Testis-Specific Gene Expression.";
RL Biochem. Biophys. Res. Commun. 263:810-815(1999).
DR EMBL; AF171929; AAF03777.1; -.
DR HSSP; P18619; 1FVL.
DR INTERPRO; IPR001590; -.
DR INTERPRO; IPR001762; -.
DR INTERPRO; IPR002870; -.
DR PFAM; PF00200; disintegrin; 1.
DR PFAM; PF01421; Reprolysin; 1.
DR PFAM; PF01562; Pep_M12B_propep; 1.
DR PRINTS; PR00289; DISINTEGRIN.
KW Integrin.
SQ SEQUENCE 820 AA; 92781 MW; B054E9FAC05E4EF5 CRC64;

Query Match 11.3%; Score 80.5; DB 4; Length 820;
Best Local Similarity 22.3%; Pred. No. 3.7;
Matches 31; Conservative 28; Mismatches 47; Indels 33; Gaps 6;

QY 18 GCKTPHLLK-----DGVCALYNEQDEAAVLEVPQHSDSLLLHCRITEADPQTSITL-- 69
DB 60 GOKHIIHKVKLLFSKHLPVFTYTDGAILLEDQPFVQNNCYH-GYVEGDPESLSVLSLT 118
QY 70 ----YSMLLQLNFEMAAAMRGCVLDELHNVR-LCFQOSLEHL-----DEASFSDIVSG 118
DB 119 CFGGFGQILQIN-----DFAYEIKPLAFSTTFEHLVYKMDSEKQFSTMRSG 165
QY 119 FIEHAAEVREYIAQLDESS 137
DB 166 FMQNEITCRMEFEIDNST 184

RESULT 10
Q9PE32 PRELIMINARY; PRT; 830 AA.
ID Q9PE32
AC Q9PE32
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE RIBONUCLEOSIDE-DIPHOSPHATE REDUCTASE ALPHA CHAIN.
GN XF1196.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OX Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9A5C;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvaranga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferto J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pasquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,
RA Vallada H., Van Sluys M.A., Verjovsky-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
```



Query Match	10.5%	Score 75;	DB 5;	Length 2053;
Best Local Similarity	20.5%;	Pred. No. 41;		
Matches	32;	Conservative	28;	Mismatches 50;
				Indels 46;
				Gaps 7;
QY	12	LOFSAGCKPIHL-----KDGV-----	CALYNEQDEEAALVLPQ-----	47
Db	1490	MRYFD--FKSPIAGETNDGLPVSICEDVFHCLMFFCRCTNQETKQALTSLGSCFVLND	1547	
QY	48	-----HSDSLLLHCRIETADPQTSITLYSLMLQLNFEMAARGCWLALDE----	LHNRLC	99
Db	1548	GYLTRSELKNLYCEI-----	LSSIANDAGFKICMNIWYLTSEMFHNKEKE	1597
QY	100	FOOSLEHLDRAFSDIVSGFIEHAAEVREYIAQLDE	135	
Db	1598	WEKOSKEDULKEMNDVSSGMASRIQL--VLEEILE	1631	

```

Query Match      10.5%; Score 75; DB 5; Length 2053;
Best Local Similarity 20.5%; Pred. No. 41;
Matches 32; Conservative 28; Mismatches 50; Indels 46; Gaps 7;

Qy 12 LQYFSAGCKTPIHL---KQGV-----CALYNEDDEAAVLEVPQ----- 47
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 1490 MRYED--FKSPALGETNDGLPVSTICEDVFHCLMFFCRCTNOETRKQALISLGSFCVLND 1547

Qy 48 ----HSSDLLHCRILIEADPQTSITLYSMLLQLPENFMAAMRGCVLALDE----LHNVRLC 99
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 1548 GYLTRSELKNLYCET-----LSSIANDAQGFKIICMNITWITYTESMFHNKKE 1597

```













Qy 109 E-ASFSDIVSG 118  
Db 210 GLAOPFDPLPG 220

RESULT 5  
OGT1\_HUMAN STANDARD; PRT; 920 AA.  
AC O15294;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE UDP-N-ACETYLGLUCOSAMINE--PEPTIDE N-ACETYLGLUCOSAMINYLTRANSFERASE 100  
DE KDA SUBUNIT (EC 2.4.1.-) (O-GLCNAC TRANSFERASE P100 SUBUNIT).  
GN OGT.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 91-110 AND 829-845.  
RX MEDLINE=97238870; PubMed=9083068;  
RA Lubas W.A., Frank D.W., Krause M., Hanover J.A.;  
RT "O-linked GlcNAc transferase is a conserved nucleocytoplasmic protein containing tetratricopeptide repeats.";  
RL J. Biol. Chem. 272:9316-9324 (1997).  
CC -!- FUNCTION: ADDITION OF NUCLEOTIDE-ACTIVATED SUGARS DIRECTLY ONTO THE POLYPEPTIDE THROUGH O-GLYCOSIDIC LINKAGE WITH THE HYDROXYL OF SERINE OR THREONINE.  
CC -!- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GLUCOSAMINE + PEPTIDE = UDP + N-ACETYL-BETA-D-GLUCOSAMINYL-PEPTIDE.  
CC -!- PATHWAY: GLYCOSYLATION.  
CC -!- SUBUNIT: HETEROTRIMER OF TWO 100 KDA AND ONE 70 KDA SUBUNITS. IT IS NOT KNOWN IF THE 70 KDA SUBUNIT IS ENCODED BY A SEPARATE GENE OR IS THE PRODUCT OF EITHER OF A PROTEOLYTIC DEGRADATION OR AN ALTERNATIVE INITIATION OF THE 100 KDA SUBUNIT (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: NUCLEAR AND CYTOPLASMIC (POSSIBLE).  
CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN PANCREAS AND TO A LESSER EXTENT IN SKELETAL MUSCLE, LIVER, BRAIN AND PLACENTA. PRESENT IN TRACE AMOUNTS IN LUNG AND HEART.  
CC -!- SIMILARITY: CONTAINS 10 TPR REPEATS.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; U77413; AAB63466.1; --  
DR MIM; 300255; --  
DR InterPro; IPR001440; --  
DR Pfam; PF00515; TPR; 8.  
KW transferase; Glycosyltransferase; Nuclear protein; Repeat; TPR repeat.  
FT REPEAT 35 64 TPR 1.  
FT REPEAT 65 98 TPR 2.  
FT REPEAT 99 132 TPR 3.  
FT REPEAT 133 166 TPR 4.  
FT REPEAT 167 200 TPR 5.  
FT REPEAT 201 234 TPR 6.  
FT REPEAT 245 268 TPR 7.  
FT REPEAT 269 302 TPR 8.  
FT REPEAT 303 336 TPR 9.  
FT REPEAT 337 347 TPR 10 (INCOMPLETE).  
FT DOMAIN 361 377 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
SQ SEQUENCE 920 AA; 103011 MW; 766BF416ABD547C4 CRC64;

Query Match 10.0%; Score 71.5; DB 1; Length 920;  
Best Local Similarity 24.4%; Pred. No. 16;  
Matches 30; Conservative 21; Mismatches 51; Indels 21; Gaps 4;

Qy 24 HLKDGVCALYNEODEAAVLEV-----POHSDSL--LLHCRRIEADPQTSITLY 70  
Db 165 HFFDAYCNLANALKEKSGVAEAEACYNATRLCPHADSNNLANIKRQGNIEEAVRLY 224  
Qy 71 SMLLOLNFEAMRGWCWALDELHNVRLCFOOSLEHLEDA-----SFSDIVSGFIEHAAE 125  
Db 225 RKALEVPEFAAHS---NLASVLQOQKLOEALMHYKEAIRISPTFADAYSNMGTKE 281  
Qy 126 VRE 128  
Db 282 MOD 284

RESULT 6  
OGT1\_RAT STANDARD; PRT; 1036 AA.  
AC P56558;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE UDP-N-ACETYLGLUCOSAMINE--PEPTIDE N-ACETYLGLUCOSAMINYLTRANSFERASE 110  
DE KDA SUBUNIT (EC 2.4.1.-) (O-GLCNAC TRANSFERASE P110 SUBUNIT).  
GN OGT.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RX STRAIN=SPRAGUE-DAWLEY; TISSUE=Liver;  
RX MEDLINE=97238869; PubMed=9083067;  
RA Kreppel L.K., Blomberg M.A., Hart G.W.;  
RT "Dynamic glycosylation of nuclear and cytosolic proteins. Cloning and characterization of a unique O-GlcNAc transferase with multiple tetratricopeptide repeats.";  
RL J. Biol. Chem. 272:9308-9315 (1997).  
CC -!- FUNCTION: ADDITION OF NUCLEOTIDE-ACTIVATED SUGARS DIRECTLY ONTO THE POLYPEPTIDE THROUGH O-GLYCOSIDIC LINKAGE WITH THE HYDROXYL OF SERINE OR THREONINE.  
CC -!- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GLUCOSAMINE + PEPTIDE = UDP + N-ACETYL-BETA-D-GLUCOSAMINYL-PEPTIDE.  
CC -!- ENZYME REGULATION: BY TYROSINE PHOSPHORYLATION AND O-GLCNAC MODIFICATIONS.  
CC -!- PATHWAY: GLYCOSYLATION.  
CC -!- SUBUNIT: HETEROTRIMER OF TWO 110 KDA AND ONE 78 KDA SUBUNITS. IT IS NOT KNOWN IF THE 78 KDA SUBUNIT IS ENCODED BY A SEPARATE GENE OR IS THE PRODUCT OF EITHER OF A PROTEOLYTIC DEGRADATION OR AN ALTERNATIVE INITIATION OF THE 110 KDA SUBUNIT.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR AND CYTOPLASMIC (POSSIBLE).  
CC -!- TISSUE SPECIFICITY: APPEARS TO BE PRESENT IN ALL TISSUES EXAMINED EXCEPT KIDNEY.  
CC -!- PTM: AUTOMODIFIED BY O-GLYCOSYLATION WITH O-GLCNAC.  
CC -!- SIMILARITY: CONTAINS 12 TPR REPEATS.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; U76557; AAC53121.1; --  
DR InterPro; IPR001440; --  
DR Pfam; PF00515; TPR; 11.  
KW transferase; Glycosyltransferase; Nuclear protein; Repeat; TPR repeat;  
KW Phosphorylation; Glycoprotein.  
FT REPEAT 79 112 TPR 1.  
FT REPEAT 113 146 TPR 2.  
FT REPEAT 147 180 TPR 3.  
FT REPEAT 181 214 TPR 4.  
FT REPEAT 215 248 TPR 5.

FT REPEAT 249 282 TPR 6.  
 FT REPEAT 283 316 TPR 7.  
 FT REPEAT 317 350 TPR 8.  
 FT REPEAT 351 384 TPR 9.  
 FT REPEAT 385 418 TPR 10.  
 FT REPEAT 419 452 TPR 11.  
 FT REPEAT 453 463 TPR 12 (INCOMPLETE).  
 FT DOMAIN 478 493 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT MOD\_RES 979 979 PHOSPHORYLATION (PROBABLE).  
 SQ SEQUENCE 1036 AA; 115605 MW; 3F057CABDD019BD6 CRC64;

Query Match 10.0%; Score 71.5; DB 1; Length 1036;  
 Best Local Similarity 24.4%; Pred. No. 18;  
 Matches 30; Conservative 21; Mismatches 51; Indels 21; Gaps 4;

OY 24 HLKDGVCALYNEQDEEAVALV-----FOHSDSL--LLHCRIEADPQTSITLY 70  
 Db 281 HFPDAYCNLANALKEGKGVAAEDCYNTALRLCPTHADSLNLANIKREOGNIEAVRLY 340  
 OY 71 SMLLQNFEMAMRGCGWLDELHNRLCFOOSLEHLDEA-----SFSDIVSGFIEHAAE 125  
 Db 341 RKALEVFPFAAAHS-----NLASVLQOQKQLQELMHYKEAIRISPTFADAYSNGMTLKE 397

OY 126 VRE 128  
 Db 398 MQD 400

RESULT 7  
 ID NEL2\_MOUSE STANDARD; PRT; 816 AA.  
 AC 061220;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE PROTEIN KINASE C-BINDING PROTEIN NELL2 PRECURSOR (NEL-LIKE PROTEIN 2)  
 DE (MEL91 PROTEIN).  
 GN NELL2 OR MEL91.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Ekins D.A., Rossi J.;  
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP TSP N-TERMINAL DOMAIN.  
 RX MEDLINE-98153258; PubMed=9480764;  
 RA Beckmann G., Hanke J., Bork P., Reich J.;  
 RT "Merging extracellular domains: fold prediction for laminin G-like and amino-terminal thrombospondin-like modules based on homology to pentraxins";  
 RL J. Mol. Biol. 275:725-730(1998).  
 CC -1- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).  
 CC -1- SIMILARITY: CONTAINS 1 TSP N-TERMINAL DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 5 WFEC DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; U59230; AAB02924.1; ALT\_INIT.  
 CC MGD; MGI:1858510; Nell2.  
 CC HSP; P00740; IIXA.  
 DR InterPro; IPR000152; -.  
 DR InterPro; IPR000561; -.

DR InterPro; IPR001007; -.  
 DR InterPro; IPR001881; -.  
 DR Pfam; PF00008; EGF; 5.  
 DR Pfam; PF00093; vwc; 2.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; 3.  
 DR PROSITE; PS00022; EGF\_1; 1.  
 DR PROSITE; PS01186; EGF\_2; 3.  
 DR PROSITE; PS01187; EGF\_CA; 3.  
 DR PROSITE; PS01208; WFEC; 2.  
 KW Glycoprotein; EGF-like domain; Repeat; Signal.  
 FT SIGNAL 1 24  
 FT CHAIN 25 816  
 FT DOMAIN 30 258  
 FT DOMAIN 272 331  
 FT DOMAIN 332 396  
 FT DOMAIN 337 439  
 FT DOMAIN 440 481  
 FT DOMAIN 482 522  
 FT DOMAIN 521 553  
 FT DOMAIN 555 601  
 FT DOMAIN 602 637  
 FT DOMAIN 638 693  
 FT DOMAIN 698 756  
 FT DOMAIN 758 813  
 FT DISULFID 401 413  
 FT DISULFID 407 422  
 FT DISULFID 424 438  
 FT DISULFID 444 457  
 FT DISULFID 451 466  
 FT DISULFID 468 480  
 FT DISULFID 486 499  
 FT DISULFID 493 508  
 FT DISULFID 510 521  
 FT DISULFID 525 535  
 FT DISULFID 529 541  
 FT DISULFID 543 552  
 FT DISULFID 559 572  
 FT DISULFID 566 581  
 FT DISULFID 583 600  
 FT DISULFID 606 619  
 FT DISULFID 613 628  
 FT DISULFID 630 636  
 FT CARBOHYD 53 53  
 FT CARBOHYD 225 225  
 FT CARBOHYD 293 293  
 FT CARBOHYD 298 298  
 FT CARBOHYD 517 517  
 FT CARBOHYD 615 615  
 FT CARBOHYD 635 635  
 SQ SEQUENCE 816 AA; 91163 MW; 5BDD0A946F87E74D CRC64;

Query Match 9.9%; Score 71; DB 1; Length 816;  
 Best Local Similarity 24.6%; Pred. No. 16;  
 Matches 33; Conservative 12; Mismatches 55; Indels 34; Gaps 4;

OY 3 SSQQRVERFLO-----YFSAGCKTPHLKDCVCALYNEQDEEAVALVPHQSDSLHL 55  
 Db 69 ASTATAERFLQKLNKHEFTILVTLUKQIHLNSGVLTSTLHLDHRYLESSGHRNEIRLH 128  
 OY 56 CRTEADPQTSITLYSMLLQNFEMAMRGCGWLDELHNRLCFOOS--LEHLD----- 108  
 Db 129 YRSQTHRPHTVEVPFYI-----LADAKWHKLSLAFSASHLILHIDCNKIY 172  
 OY 109 -----EASFSDIVSG 118  
 Db 173 ERVVEPFTDLALG 186

RESULT 8  
 POLG\_EC09B  
 ID POLG\_EC09B STANDARD; PRT; 2203 AA.  
 AC Q66577;





Query Match 9.7%; Score 69; DB 1; Length 451;  
Best Local Similarity 22.7%; Pred. No. 12;  
Matches 29; Conservative 30; Mismatches 33; Indels 36; Gaps 7;  
20 KPTIHLKGVGKALYNQDEBAAVLEVPQHSDDLHLHC-----RIEADPQTSITLIYS 71

Db	231	KLP-HFEEGAVTVODLSAQWNAATLLEPKNEWILDACAAPGGKTHHLELAQANV----	285
QY	72	MLQLNFEMAAMRGWCWLAD-ELNRNLCRFOOSLEHLDEAFSDIVSGFIEHAAEVREYI	130
Db	286	-----IALDVESHRLKR-VEENLERLNQQAAT--VVCG---DASKPDEWL	323
QY	131	AQDESSA 138	
Db	324	AEIGKSAA 331	
RESULT 13			
MERL_HUMAN			
ID	MERL_HUMAN	STANDARD;	PRT; 595 AA.
AC	P35240;		
DT	01-FEB-1994 (Rel. 28, Created)		
DT	01-FEB-1994 (Rel. 28, Last sequence update)		
DT	01-OCT-2000 (Rel. 40, Last annotation update)		
DE	MERLIN (SCHWANNOMIN) (NEUROFIBROMIN 2).		
DE	NF2 OR SCH.		
GN	Homo sapiens (Human).		
OS	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
QY	[1]		
RN	SEQUENCE FROM N.A.		
RX	MEDLINE=93201601; PubMed=8453669;		
RX	Trofater J.A., Maccollin M.M., Rutter J.L., Murrell J.R., Duyao M.P.,		
RA	Parry D.N., Eldridge R., Kley N., Menon A.G., Pulaski K., Haase V.H.,		
RA	Amrose C.M., Munroe D., Bove C., Haines J.L., Martuza R.L.,		
RA	Macdonald M.E., Seizinger B.R., Short M.P., Buckler A.J.,		
RA	Gusella J.F.,		
RT	"A novel moesin-, ezrin-, radixin-like gene is a candidate for the		
RT	neurofibromatosis 2 tumor suppressor."		
RL	Cell 72:791-800(1993).		
RN	[2]		
RN	SEQUENCE FROM N.A.		
RX	MEDLINE=93281181; PubMed=8379998;		
RA	Rouleau G.A., Merel P., Lutchman M., Sanson M., Zucman J.,		
RA	Marineau C., Hoang-Xuan K., Demczuk S., Desmaze C., Plougastel B.,		
RA	Pulst S., Lenoir G., Bijsma E., Fashold R., Dumanski J., de Jong P.,		
RA	Parry D., Eldridge R., Aurias A., Delattre O., Thomas G.;		
RT	"Alteration in a new gene encoding a putative membrane-organizing		
RT	protein causes neuro-fibromatosis type 2."		
RL	Nature 363:515-521(1993).		
RN	[3]		
RP	REVIEW.		
RA	Marineau C., Merel P., Rouleau G.A., Thomas G.;		
RT	"The gene of neurofibromatosis type 2;"		
RL	Medicine/Sciences 11:35-42(1995).		
RN	[4]		
RP	VARIANT NF2 TYR-220.		
RX	MEDLINE=94047514; PubMed=8230593;		
RA	Maccollin M., Mohny T., Trofatter J., Wertelecki W., Ramesh V.,		
RA	Gusella J.F.;		
RT	"DNA diagnosis of neurofibromatosis 2. Altered coding sequence of the		
RT	merlin tumor suppressor in an extended pedigree."		
RL	JAMA 270:2316-2320(1993).		
RN	[5]		
RP	VARIANT NF2 PHE-96 DEL.		
RX	MEDLINE=94311312; PubMed=7913580;		
RA	Maccollin M., Ramesh V., Jacoby L.B., Louis D.N., Rubio M.-P.,		
RA	Pulaski K., Trofatter J.A., Short M.P., Bove C., Eldridge R.,		
RA	Parry D.M., Gusella J.F.;		
RT	"Mutation analysis of patients with neurofibromatosis 2;"		
RL	Am. J. Hum. Genet. 55:314-320(1994).		
RN	[6]		
RP	VARIANT ARG-46.		
RX	MEDLINE=94272478; PubMed=8004107;		
RA	Irving R.M., Moffat D.A., Hardy D.G., Barton D.E., Xuereb J.H.,		
RA	Maher E.R.;		
RT	"Somatic NF2 gene mutations in familial and non-familial vestibular		

schwannoma.";  
Hum. Mol. Genet. 3:347-350(1994).  
[7]  
VARIANTS MET-219 AND CYS-418  
MEDLINE=94282032; PubMed=8012353;  
RA Jacoby L.B., Maccollin M., Louis D.N., Mohney T., Rubio M.P.,  
RA Pulaski K., Trofatter J.A., Kley N., Seizinger B., Ramesh V.,  
RA Gusella J.F.;  
RT "Exon scanning for mutation of the NF2 gene in schwannomas.";  
RL Hum. Mol. Genet. 3:413-419(1994).  
[8]  
VARIANTS NF2 SER-62; GLY-106 AND MET-352.  
MEDLINE=94362682; PubMed=8081368;  
RA Bourn D., Carter S.A., Mason S., Gareth D., Evans R., Strachan T.;  
RT "Germline mutations in the neurofibromatosis type 2 tumour suppressor  
gene.";  
RL Hum. Mol. Genet. 3:813-816(1994).  
[9]  
VARIANT GLY-79.  
MEDLINE=95038740; PubMed=7951231;  
RA Sainz J., Huynh D.P., Figueroa K., Ragge N.K., Baser M.E., Pulst S.M.;  
RT "Mutations of the neurofibromatosis type 2 gene and lack of the gene  
product in vestibular schwannomas.";  
RL Hum. Mol. Genet. 3:885-891(1994).  
[10]  
VARIANTS PHE-273 AND ILE-364.  
MEDLINE=94214500; PubMed=8162073;  
RA Bianchi A.B., Hara T., Ramesh V., Gao J., Klein Szanto A.J., Morin F.,  
RA Menon A.G., Trofatter J.A., Gusella J.F., Seizinger B.R., Kley N.;  
RT "Mutations in transcript isoforms of the neurofibromatosis 2 gene in  
multiple human tumour types.";  
RL Nat. Genet. 6:185-192(1994).  
[11]  
VARIANT NF2 PRO-535.  
MEDLINE=95395825; PubMed=7666400;  
RA Evans D.G.R., Bourn D., Wallace A., Ramsden R.T., Mitchell J.D.,  
RA Strachan T.;  
RT "Diagnostic issues in a family with late onset type 2  
neurofibromatosis.";  
RL J. Med. Genet. 32:470-474(1995).  
[12]  
VARIANT NF2 PRO-538.  
MEDLINE=96163499; PubMed=8566958;  
RA Kluewe L., Mautner V.-F.;  
RT "A missense mutation in the NF2 gene results in moderate and mild  
clinical phenotypes of neurofibromatosis type 2.";  
RL Hum. Genet. 97:224-227(1996).  
[13]  
VARIANTS F-96 DEL; I-117; F-119 DEL; V-122--E-129 DEL AND F-339.  
MEDLINE=96209910; PubMed=8655144;  
RA de Vitlis L.R., Tedde A., Vitelli F., Ammannati F., Mennonna P.,  
RA Bigozzi U., Montali E., Papi L.;  
RT "Screening for mutations in the neurofibromatosis type 2 (NF2) gene  
in sporadic meningiomas.";  
RL Hum. Genet. 97:632-637(1996).  
[14]  
VARIANT NF2 ARG-234.  
MEDLINE=99192373; PubMed=10090912;  
RA Baser M.E., Kluewe L., Mautner V.-F.;  
RT "Germline NF2 mutations and disease severity in neurofibromatosis  
type 2 patients with retinal abnormalities.";  
RL Am. J. Hum. Genet. 64:1230-1233(1999).  
CC -1- FUNCTION: PROBABLY ACTS AS A MEMBRANE STABILIZING PROTEIN.  
CC -1- TISSUE SPECIFICITY: IN FETAL BRAIN; IN KIDNEY, LUNG, BREAST,  
CC AND OVARY.  
CC -1- DISEASE: NEUROFIBROMATOSIS 2 (NF2) OR CENTRAL NEUROFIBROMATOSIS IS  
CC A GENETIC DISORDER CHARACTERIZED BY BILATERAL VESTIBULAR  
CC SCHWANNOMAS (FORMERLY CALLED ACOUSTIC NEUROMAS), SCHWANNOMAS OF  
CC OTHER CRANIAL AND PERIPHERAL NERVES, MENINGIOMAS, AND EPENDYOMAS.  
CC IT IS INHERITED IN AN AUTOSOMAL DOMINANT FASHION WITH FULL  
CC PENETRANCE. AFFECTED INDIVIDUALS GENERALLY DEVELOP SYMPTOMS OF  
CC EIGHTH NERVE DYSFUNCTION IN EARLY ADULTHOOD, INCLUDING DEAFNESS  
CC AND BALANCE DISORDER. ALTHOUGH THE TUMORS OF NF2 ARE



QY 33 YNEQDEAAVLEVPQHSLLHLCRIIE---ADPOTSITLYSMLLQ---NFEMAAMRG 85  
 DB 101 YPENAEELVQBITQHLFFLVQKQILDEKVCYCPPEASVLLASVAYQAKYGDYDPSVHKR 160  
 QY 86 CWLADELHNVRL--CFQOSLEHLEDEASFDIVSGFIEHAAVEVRE 128  
 DB 161 GFLAQEELLPKRVINLYQMTPEWEE---ITAWYAEHRGRARD 201

RESULT 14  
 MERL\_MOUSE STANDARD; PRT; 596 AA.  
 AC P46662:  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE MERLIN (SCHWANNOMIN) (NEUROFIBROMIN 2).  
 GN NF2 OR NF-2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=94282031; PubMed=8012352;  
 RA Haase V.H., Trofatter J.A., Maccollin M., Tarttelin E., Guseella J.F.,  
 RA Ramesh V.;  
 RT "The murine NF2 homologue encodes a highly conserved merlin protein  
 RT with alternative forms.";  
 RL Hum. Mol. Genet. 3:407-411(1994).  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=95072570; PubMed=7981675;  
 RA Huynh D.P., Nechiporuk T., Pulst S.M.;  
 RT "Alternative transcripts in the mouse neurofibromatosis type 2 (NF2)  
 RT gene are conserved and code for schwannomins with distinct C-terminal  
 RT domains.";  
 RL Hum. Mol. Genet. 3:1075-1079(1994).  
 RN [3]  
 RN SEQUENCE OF 400-596 FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=94375073; PubMed=8088840;  
 RA Claudio J.O., Marneau C., Rouleau G.A.;  
 RT "The mouse neurofibromatosis type 2 gene maps to chromosome 11.";  
 RL Genomics 21:437-439(1994).  
 CC -!- FUNCTION: PROBABLY ACTS AS A MEMBRANE STABILIZING PROTEIN.  
 CC -!- ALTERNATIVE PRODUCTS: AT LEAST 3 ISOFORMS ARE PRODUCED BY  
 CC ALTERNATIVE SPLICING.  
 CC -!- SIMILARITY: CONTAINS A DOMAIN FOUND IN BAND 4.1, EZRIN, MOESIN,  
 CC RADIXIN, AND TALIN.

-----  
 THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 EMBL; X74671; CAA52737.1; -  
 EMBL; L27105; AAA39807.1; -  
 EMBL; L27090; AAA63648.1; -  
 EMBL; L28176; AAA39808.1; -  
 EMBL; X75759; CAA53386.1; -  
 DR MGD; MGI:97307; NF2.  
 DR InterPro; IPR000299; -  
 DR InterPro; IPR000798; -  
 DR Pfam; PF00373; Band\_41; 1.  
 DR Pfam; PF00769; ERW; 1.  
 DR PRINTS; PR00935; BAND41.  
 DR PROSITE; PS00660; BAND\_41\_1; 1.  
 DR PROSITE; PS00661; BAND\_41\_2; 1.

DR PROSITE; PS00057; BAND\_41\_3; 1.  
 KW Structural protein; Cytoskeleton; Anti-oncogene; Alternative splicing.  
 FT DOMAIN 74 241  
 FT VARSPLIC 581 596  
 FT LTLSAKSRVAFEEEL -> PQAQRRPICI (IN  
 FT SECOND ISOFORM).  
 FT I -> T (IN REF. 2).  
 FT CONFLICT 475 475  
 FT CONFLICT 554 554  
 FT CONFLICT 570 570  
 FT CONFLICT G -> A (IN REF. 2).  
 FT SEQUENCE 596 AA; 69691 MW; 8D06E3B07D2C5851 CRC64;

Query Match 9.6%; Score 68.5; DB 1; Length 596;  
 Best Local Similarity 25.7%; Pred. No. 19;  
 Matches 27; Conservative 22; Mismatches 43; Indels 13; Gaps 4;  
 QY 33 YNEQDEAAVLEVPQHSLLHLCRIIE---ADPOTSITLYSMLLQ---NFEMAAMRG 85  
 DB 101 YPENAEELVQBITQHLFFLVQKQILDEKVCYCPPEASVLLASVAYQAKYGDYDPSVHKR 160  
 QY 86 CWLADELHNVRL--CFQOSLEHLEDEASFDIVSGFIEHAAVEVRE 128  
 DB 161 GFLAQEELLPKRVINLYQMTPEWEE---ITAWYAEHRGRARD 201

RESULT 15  
 ARDE\_CHLMU STANDARD; PRT; 478 AA.  
 AC P56961:  
 DT 01-OCT-2000 (Rel. 40, Created)  
 DT 01-OCT-2000 (Rel. 40, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE SHIKIMATE BIOSYNTHESIS PROTEIN ARODE [INCLUDES: 3-DEHYDROQUINATE  
 DE DEHYDRATASE (EC 4.2.1.10) (3-DEHYDROQUINASE) (TYPE I DHOASE)];  
 DE SHIKIMATE 5-DEHYDROGENASE (EC 1.1.1.25)].  
 GN AROE OR AROD/E OR TC0649.  
 OS Chlamydia muridarum.  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_TaxID=83560;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=MOPN / NIGG;  
 RX MEDLINE=20150255; PubMed=10684935;  
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
 RA White O., Hickey E.K., Peterson J., Utterback T., Berry K.,  
 RA Bass S., Linher K., Weidman J., Khouri H., Craven B., Bowman C.,  
 RA Dodson R., Gwinn M., Nelson W., Deboy R., Kolonay J., McClarty G.,  
 RA Salzberg S.L., Eisen J., Fraser C.M.;  
 RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia  
 RT pneumoniae AR39.";  
 RL Nucleic Acids Res. 28:1397-1406(2000).

CC -!- FUNCTION: BIFUNCTIONAL ENZYME THAT CATALYZES TWO SEQUENTIAL STEPS  
 CC OF THE AROMATIC AMINO ACIDS BIOSYNTHETIC PATHWAY. THE FIRST  
 CC REACTION IS CATALYZED BY THE 3-DEHYDROQUINATE DEHYDRATASE, CODED  
 CC BY THE AROD DOMAIN; THE SECOND REACTION IS CATALYZED BY THE  
 CC SHIKIMATE 5-DEHYDROGENASE, CODED BY THE AROE DOMAIN.  
 CC -!- CATALYTIC ACTIVITY: 3-DEHYDROQUINATE = 3-DEHYDROSHIKIMATE + H(2)O.  
 CC -!- CATALYTIC ACTIVITY: SHIKIMATE + NADP(+) = 5-DEHYDROSHIKIMATE +  
 CC NADPH.  
 CC -!- PATHWAY: CATALYZES THE THIRD (AROD) AND FOURTH (AROE) STEPS IN THE  
 CC BIOSYNTHESIS FROM CHORISMATE OF THE AROMATIC AMINO ACIDS (THE  
 CC SHIKIMATE PATHWAY).  
 CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE TYPE-I 3-  
 CC DEHYDROQUINASE FAMILY.  
 CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE SHIKIMATE  
 CC DEHYDROGENASE FAMILY.  
 CC -----  
 THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----

```

CC -----
DR EMBL: AE002332; AAF39475.1; -.
DR TIGR: TC0649; -.
DR InterPro: IPR001381; -.
DR InterPro: IPR002907; -.
DR Pfam: PF01487; DHQuinase_I; 1.
DR Pfam: PF01488; Shikimate_DH; 1.
DR PROSITE: PS01028; DEHYDROQUINASE_I; FALSE_NEG.
KW Aromatic amino acid biosynthesis; Lyase; Oxidoreductase; NADP;
KW Multifunctional enzyme.
FT DOMAIN 1 208 3-DEHYDROQUINATE DEHYDRATASE.
FT DOMAIN 209 478 SHIKIMATE 5-DEHYDROGENASE.
FT ACT_SITE 110 110 BY SIMILARITY.
FT ACT_SITE 133 133 FORMS A SCHIFF-BASE INTERMEDIATE (BY
FT ACT_SITE 133 133 SIMILARITY).
SQ SEQUENCE 478 AA; 53210 MW; 199EEF5251205AB9 CRC64;

```

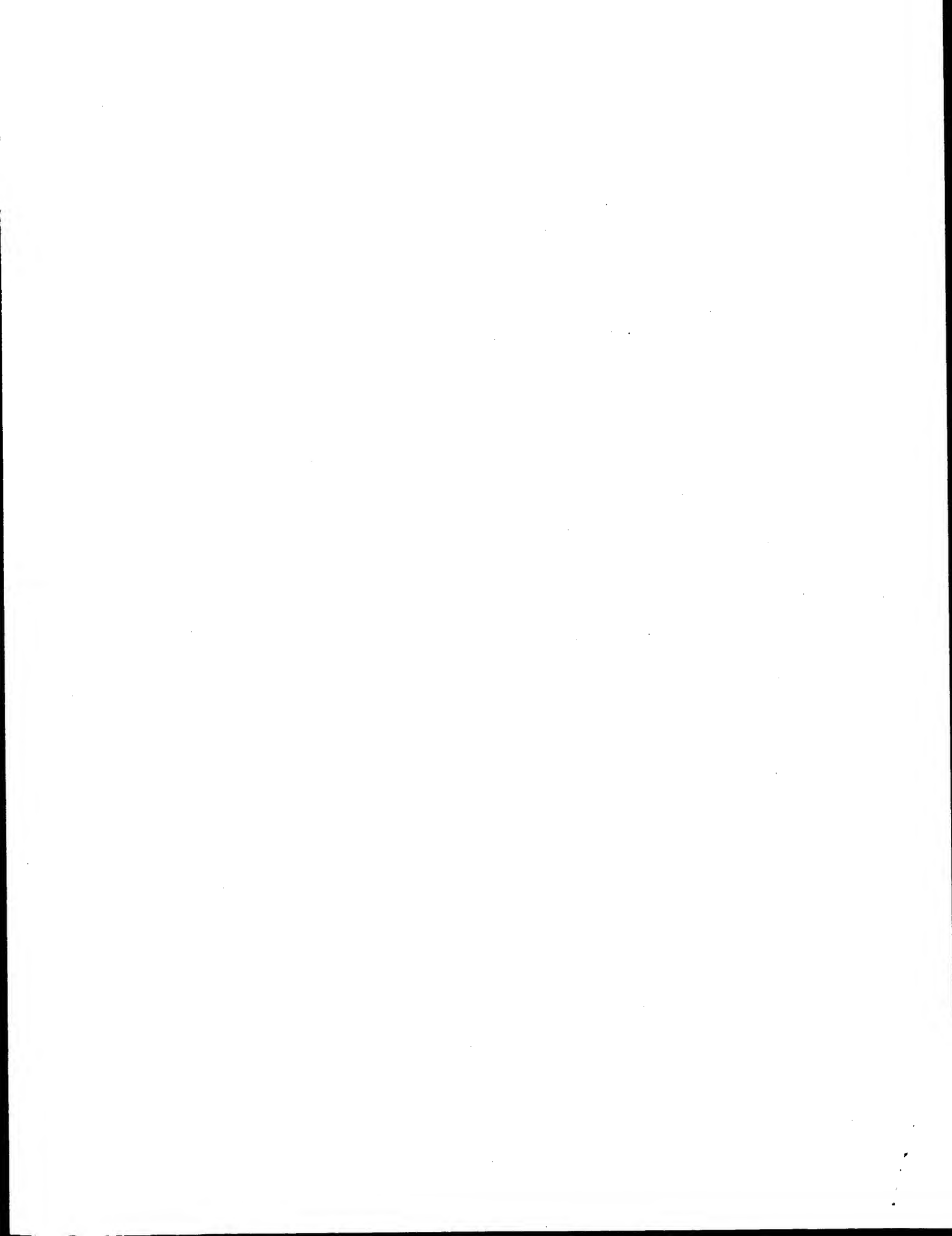
Query Match 9.5%; Score 68; DB 1; Length 478;  
 Best Local Similarity 26.4%; Pred. No. 17;  
 Matches 32; Conservative 20; Mismatches 35; Indels 34; Gaps 9;

```

QY 18 GC-----KTPHILKGVCALYNEQDEEAVALVPEQHSDSLHLHCRIEADPQTSI---T 68
Db 353 GCELLIFNRTKIH-AEAIASRYQAQAFD--IKDLPLHSVSLIINCL-----PPSSIIPA 404
QY 69 LYSMLLQLN-----FEMAAMRG-----WLDELHNVRLCF-QQSLEHLDEASF 112
Db 405 LAPLIVDINTPLPKHNSFTQYARLKGCSIYGHMFQAQALLQFRLWFPPTHFNHL-EKNF 463
QY 113 S 113
Db 464 S 464

```

Search completed: June 5, 2001, 18:23:09  
 Job time: 365 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2001, 18:18:11 ; Search time 54.28 Seconds  
(without alignments)  
175.985 Million cell updates/sec

Title: US-09-596-784-4  
Perfect score: 715  
Sequence: 1 MTSSQQRVERFLQYFSAGCK.....IEHAAEVREYIAQLDESSAA 139

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 68722935 residues  
Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_67:.\*  
1: pir1:.\*  
2: pir2:.\*  
3: pir3:.\*  
4: pir4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	715	100.0	139	2 T18449	pathogenicity factor
2	273.5	38.3	129	2 T30333	avirulence protein
3	79	11.0	789	2 A70868	probable plsB2 - M
4	78.5	11.0	830	2 C82710	ribonucleoside-dip
5	77.5	10.8	775	2 T45238	probable transfera
6	75.5	10.6	402	2 T19390	hypothetical prote
7	75.5	10.6	966	2 S25365	CyC8 protein - yea
8	73.5	10.3	984	1 A34076	protein-tyrosine k
9	71.5	10.0	1036	2 T31673	N-acetylglucosamin
10	71.5	10.0	2484	2 T26216	hypothetical prote
11	71.5	10.0	2607	2 T26215	hypothetical prote
12	70.5	9.9	1048	2 T31653	hypothetical prote
13	69.5	9.7	291	2 T23051	hypothetical prote
14	69.5	9.7	393	2 S76366	hypothetical prote
15	69.5	9.7	2183	1 ZLN2MY	genome polyprotein
16	69.5	9.7	2183	1 G48556	pts iia protein +
17	69	9.7	225	2 F72124	membrane-associate
18	69	9.7	374	2 A42264	hypothetical prote
19	69	9.7	451	1 F64155	probable outer den
20	69	9.7	610	2 T02298	outer dense fiber
21	69	9.7	638	2 T09400	hypothetical prote
22	68.5	9.6	165	2 C72625	conserved hypothe
23	68.5	9.6	167	2 H81717	hypothetical prote
24	68.5	9.6	173	2 S75578	hypothetical prote
25	68.5	9.6	352	2 T44368	gas-vesicle operon
26	68.5	9.6	591	2 I54368	merlin protein - m
27	68.5	9.6	595	2 S33809	neurofibromin 2 -
28	68.5	9.6	596	2 I68664	merlin - mouse
29	68.5	9.6	621	2 T06717	hypothetical prote

30	68.5	9.6	794	2 T46073	hypothetical prote
31	68	9.5	178	2 T39722	ORF12 - Agrobacter
32	68	9.5	478	2 G81679	3-dehydroquinat d
33	68	9.5	638	2 T03791	outer dense fiber
34	68	9.5	923	2 T24712	hypothetical prote
35	68	9.5	1119	2 B70126	surface-located me
36	68	9.5	1144	2 A54810	TMV resistance pro
37	67.5	9.4	167	2 A71563	hypothetical prote
38	67.5	9.4	1123	2 T51517	hypothetical prote
39	67.5	9.4	1141	2 T20611	telomerase reverse
40	67.5	9.4	1196	2 S65245	translation elonga
41	67.5	9.4	1527	2 JE0336	canalicular multis
42	67	9.4	578	2 T15736	hypothetical prote
43	67	9.4	746	2 S66963	hypothetical prote
44	67	9.4	748	2 G83927	alpha-galactosidas
45	67	9.4	921	2 D82513	malt regulatory pr

## ALIGNMENTS

RESULT 1  
T18449  
pathogenicity factor DspB - Erwinia amylovora  
C:Species: Erwinia amylovora  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 20-Jun-2000  
C:Accession: T18449  
R:Gaudriault, S.  
submitted to the EMBL Data Library, May 1998  
A:Reference number: Z18936  
A:Accession: T18449  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-139 <GAU>  
A:Cross-references: EMBL:Y13831; PIDN:CAA74157.1  
A:Experimental source: strain CFBP1430; specific host Pommoideae  
C:Genetics:  
A:Note: dspB  
C:Function:  
A:Description: involved in pathogenicity

Query Match		100.0%	Score 715;	DB 2;	Length 139;
Best Local Similarity		100.0%	Pred. No. 3.8e-64;		
Matches 139;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MTSSQQRVERFLQYFSAGCKTP	PIHLKDGVCALYNEQDEEAAVLEVPQHSDSLHCRITTE	60	
Db	1	MTSSQQRVERFLQYFSAGCKTP	PIHLKDGVCALYNEQDEEAAVLEVPQHSDSLHCRITTE	60	
QY	61	ADPOTSITLYSMLQLNFEMAMRCGLALDELHNVRLCFQOSLEHLDSEASFSDIVSGFT	120		
Db	61	ADPOTSITLYSMLQLNFEMAMRCGLALDELHNVRLCFQOSLEHLDSEASFSDIVSGFT	120		
QY	121	EHAEEVREYIAQLDESSAA	139		
Db	121	EHAEEVREYIAQLDESSAA	139		

RESULT 2  
T30333  
avirulence protein - Pseudomonas syringae  
C:Species: Pseudomonas syringae  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999  
C:Accession: T30333  
R:Bogdanove, A.J.; Kim, J.P.; Wei, Z.; Kolchinsky, P.; Charkowski, A.O.; Conlin, A.K.  
Proc. Natl. Acad. Sci. U.S.A. 95, 1325-1330, 1998  
A:Title: Homology and functional similarity of an hrp-linked pathogenicity locus, dsf  
A:Reference number: Z20825; MUID:98115919  
A:Accession: T30333  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-129 <BOG>

A:Cross-references: EMBL:U97505; NID:32978502; PID:32978504; PIDN:AAC06135.1

C:Genetics: avif

Query Match 38.3%; Score 273.5; DB 2; Length 129;  
Best Local Similarity 43.1%; Pred. No. 3e-20;  
Matches 56; Conservative 27; Mismatches 44; Indels 3; Gaps 3;

Qy 1 MTSSQORVERFLOYFSAGCKTPIHLKDGVCALYNEDEEAAVLEVPQHSLSLLHCRITE 60

Db 1 MKTSPDFARFNSLGAQIGTSLTONGVCALYDGGNNEAIIIEPEHSEWIFHCIGR 60

Qy 61 ADPQTSITLYSMLQLNFEMAMRGCLALDELHNRLCFQOQSLHLEDEASFDIVSGFI 120

Db 61 C-PERAPDLL-RLLSLNFDVRLHGCWFVADQ-GDVRCLCAORELASLDEPAFCVTRGEI 117

Qy 121 EHAAEVREYI 130

Db 118 SQAREARAF 127

# RESULT 3

A70868 probable plSB2 - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
C:Accession: A70868

R:Colo, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987

A:Accession: A70868

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1789 <COL>

A:Cross-references: GB:AL021246; GB:AL123456; NID:3261507; PIDN:CAA16059.1; PID:el23760

C:Genetics:

A:Gene: plSB2

Query Match 11.0%; Score 79; DB 2; Length 789;  
Best Local Similarity 24.3%; Pred. No. 6.1;  
Matches 34; Conservative 23; Mismatches 51; Indels 32; Gaps 4;

Qy 23 IHLKDGVCALYNEDEEAAVLEVPQHSLSLLHCRITADPQTSITLYSMLQLNFEMAA 82

Db 445 VSMRQYLCAPHGELTQDPAKRLALQKMSFEVAMRILQATPVATGLVSALL-----LT 498

Qy 83 MRGCLALDELHNRLCFQOQSLHLEDEASFDIVSGF-----EASFSDIVSGF----- 119

Db 499 TRGTALTLDQLHHT---LQDSLIDYLERKQSPVSTSLRLRSREGVRAAADALSNHGPVTR 555

Qy 120 IEHAAEVREYIAQLDESSAA 139

Db 556 VDSGREPVWYIAPDDEHAAA 575

# RESULT 4

C82710 ribonucleoside-diphosphate reductase alpha chain Xf1196 [imported] - Xylella fastidiosa

C:Species: Xylella fastidiosa

C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Sep-2000

C:Accession: C82710

R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing

A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A:Reference number: A82515; MUID:20365717

A:Note: for a complete list of authors see reference number A59328 below

A:Accession: C82710

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-830 <SIM>

A:Cross-references: GB:AE003954; GB:AE003849; NID:99106165; PIDN:AAF84006.1; GSPDB:GN

A:Experimental source: strain 9a5c

R:Simson, A.J.G.; Reinach, F.C.; Artuda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carreras-Neto, E.; Docena, C.; El-Dorriy, H.; Facincani, A.P.; Ferreira, A.J.S.

submitted to Genbank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; La J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; La Chado, M.A.; Madeira, A.M.B.N.; Marinho, C.L.; Marques, M.V.; Martins, A.; Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, R.; Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Silva; A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silva M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.

A:Reference number: A59328

A:Contents: annotation

C:Genetics:

A:Gene: XF1196

C:Superfamily: herpesvirus ribonucleoside-diphosphate reductase large chain

Query Match 11.0%; Score 78.5; DB 2; Length 830;  
Best Local Similarity 25.6%; Pred. No. 7.3;  
Matches 33; Conservative 21; Mismatches 46; Indels 29; Gaps 6;

Qy 17 AGCKTPHILKDGVCALYNEDEEAAVLEVPQHSLSLLHCRITADPQTSITLYSMLQ 75

Db 22 AGSVFPVALS-----VFHPDEPED-----EVPPIPTAMHMTVAEEANVSWITKEAGNRR 73

Qy 76 LNFEMAMRGCLALDELHNRLCFQOQSLHLEDEASFDIVSGFIEHAAEV----- 126

Db 74 MFEDCARLE---CAIDTIH-----QEPQLDVAEYKRAVFGVERKDSVNDLVDLL 123

Qy 127 -REYIAQLD 134

Db 124 TREAEARVD 132

# RESULT 5

T45238 probable transferase [imported] - Mycobacterium leprae

C:Species: Mycobacterium leprae

C:Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 21-Jan-2000

C:Accession: T45238

R:James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL data library, March 1999

A:Reference number: Z22949

A:Accession: T45238

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1775 <TAM>

A:Cross-references: EMBL:AL049913; PIDN:CAB43153.1

A:Experimental source: cosmid B1610

C:Genetics:

A:Note: plsB

Query Match 10.8%; Score 77.5; DB 2; Length 775;  
Best Local Similarity 25.8%; Pred. No. 8.5;  
Matches 34; Conservative 16; Mismatches 33; Indels 49; Gaps 5;

Qy 11 FLQYFSAGCKTPIHLKDGVCALYN-----EGVAMLYSPIKAQGRNYKIVRFPPEAVSMRQYLGAHPGALVQ 36

Db 397 YAAARGEKTP----EGVAMLYSPIKAQGRNYKIVRFPPEAVSMRQYLGAHPGALVQ 452

Qy 37 DEEAVALVEVPQHSLSLLHCRITADPQTSITLYSMLQLNFEMAMRGCLALDELHN 96

Db 453 DQDAKRLALQKMSFEVA--WRILCATPVATGLVSALL-----LTRGVALTLDQLHHT 504



Wed Jun 6 10:19:17 2001

A:Map position: 7q32-7q36  
C:Superfamily: protein-tyrosine kinase, receptor type eph; fibronectin type III repeat  
C:Keywords: Anp; autophosphorylation; glycoprotein; kinase-related transforming protein;  
F;1-23/Domain: signal sequence #status predicted <SIG>  
F;1-984/Product: protein-tyrosine kinase receptor type eph 1 #status predicted <MAT>  
F;548-568/Domain: transmembrane #status predicted <TM>  
F;630-895/Region: protein kinase homology <KIN>  
F;638-646/Region: protein kinase ATP-binding motif  
F;918-984/Domain: SAM homology <SAM>  
F;59,338,414,478/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 10.0%; Score 73.5; DB 1; Length 984;  
Best Local Similarity 27.5%; Pred. No. 28;  
Matches 36; Conservative 20; Mismatches 52; Indels 23; Gaps 6;

QY 7 RVERFLOYSAGCKT-----PIHLKDGVCALYNEQDEEAALVLEVPQHSDSLLLHCRRIIE 60  
Db 94 RVHVELQTVTRDCKSPGAGPLGCKETFNLLYMSDQDVGi-----QLRRPLFKQVTTVA 149  
QY 61 ADPQTSI-TLYSMLLQNLNFMAM-----RGWLALDE-----LHNRLCFOOSLEHLD 108  
Db 150 ADOSFTIRDLASGVKLVNVERCSLGRLTRGLYLAFHPNPGACVALYSVRVYQRCPETLN 209  
QY 109 E-ASFSDIVSG 118  
Db 210 GLAQFPDPLPG 220

RESULT 9  
T31673  
N-acetylglucosaminyltransferases (EC 2.4.1.-), chain p110 - rat  
N:Alternate names: O-GlcNAc transferase  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 07-Dec-1999  
C:Accession: T31673  
R:Krappe, L.K.; Blomberg, M.A.; Hart, G.W.  
J. Biol. Chem. 272, 9308-9315, 1997  
A:Title: Dynamic glycosylation of nuclear and cytosolic proteins. Cloning and characteri  
A:Reference number: Z21056; MUID:97238869  
A:Accession: T31673  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1036 <KRE>  
A:Cross-references: EMBL:U76557; NID:g1931578; PID:g1931579; PIDN:AAC53121.1  
A:Experimental source: strain Sprague Dawley; liver  
C:Genetics:  
A:Gene: OGT  
C:Keywords: glycosyltransferase; hexosyltransferase; tandem repeat

Query Match 10.0%; Score 71.5; DB 2; Length 1036;  
Best Local Similarity 24.4%; Pred. No. 47;  
Matches 30; Conservative 21; Mismatches 51; Indels 21; Gaps 4;

QY 24 HLKDGVCALYNEQDEEAALVLEVPQHSDSL--LLHCRRIIEADPQTSITLY 70  
Db 281 HFPDVCNTANALKEKGSVAEDCYNTALRLCPTHADSLNLANIRKQGNIEAVRLY 340  
QY 71 SMLLQNLNFMAMRGCVLALDELHNRLCFOOSLEHLEA-----SFSDIVSGFIEHAAE 125  
Db 341 RKALEVPEFAAAH---NLASVLQOQKQLQALMHYKEAIRISPTTFADAYSNNMGTLKE 397  
QY 126 VRE 128  
Db 398 MQD 400

RESULT 10  
T26216  
hypothetical protein W06A7.3c - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T26216  
R:Ainscough, R.  
submitted to the EMBL Data Library, August 1996  
A:Reference number: Z20173  
A:Accession: T26216  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2484 <WIL>  
A:Cross-references: EMBL:Z78066; PIDN:CAB51467.1; GSPDB:GN00023; CESP:W06A7.3c  
A:Experimental source: clone W06A7  
C:Genetics:  
A:Gene: CESP:W06A7.3c  
A:Map position: 5  
A:Introns: 1827/1; 1866/3; 1963/3; 1990/3; 2262/1; 2289/1; 2339/2; 2463/2

Query Match 10.0%; Score 71.5; DB 2; Length 2484;  
Best Local Similarity 24.2%; Pred. No. 1.4e+02;  
Matches 31; Conservative 19; Mismatches 43; Indels 35; Gaps 6;

QY 18 GCKTPHILKDGVCALYNEQDEEA-----AVLEVPHQSDSLLLHCRRIIEADPOT 65  
Db 1491 GCITDVASD-----VNEQDEESTLKILKVVVPSSELSLELDFNDPKVIHVPIPLMEPAT 1545  
QY 66 SITLYSMLLQNLNFMAMRGCVLALDELHNRLCFOOSLEHLEAASFSDIVSGFIEHAAE 125  
Db 1546 -----MYLE---EMVE---WIADAVKEV-----SEMEVVTSEISEMAPQVSESTCP 1587

QY 126 VREYIAOL 133  
Db 1588 IPEPLADL 1595

RESULT 11  
T26215  
hypothetical protein W06A7.3a - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T26215  
R:Ainscough, R.  
submitted to the EMBL Data Library, August 1996  
A:Reference number: Z20173  
A:Accession: T26215  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2607 <WIL>  
A:Cross-references: EMBL:Z78066; PIDN:CAB01522.2; GSPDB:GN00023; CESP:W06A7.3a  
A:Experimental source: clone W06A7  
C:Genetics:  
A:Gene: CESP:W06A7.3a  
A:Map position: 5  
A:Introns: 1827/1; 1866/3; 1963/3; 1990/3; 2262/1; 2289/1; 2412/1; 2462/2; 2586/2

Query Match 10.0%; Score 71.5; DB 2; Length 2607;  
Best Local Similarity 24.2%; Pred. No. 1.4e+02;  
Matches 31; Conservative 19; Mismatches 43; Indels 35; Gaps 6;

QY 18 GCKTPHILKDGVCALYNEQDEEA-----AVLEVPHQSDSLLLHCRRIIEADPOT 65  
Db 1491 GCITDVASD-----VNEQDEESTLKILKVVVPSSELSLELDFNDPKVIHVPIPLMEPAT 1545  
QY 66 SITLYSMLLQNLNFMAMRGCVLALDELHNRLCFOOSLEHLEAASFSDIVSGFIEHAAE 125  
Db 1546 -----MYLE---EMVE---WIADAVKEV-----SEMEVVTSEISEMAPQVSESTCP 1587

QY 126 VREYIAOL 133  
Db 1588 IPEPLADL 1595

RESULT 12  
T31653



hypothetical protein Y57A10A.ff - *Caenorhabditis elegans*  
C:Species: *Caenorhabditis elegans*  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T31653  
R:Smeye, R.  
submitted to the EMBL Data Library, September 1999  
A:Reference number: Z21048  
A:Accession: T31653  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1048 <ML>  
A:Cross-references: EMBL:AL117195; NID:e1549729; PIDN:CAB55036.1; CESP:Y57A10A.ff  
A:Experimental source: clone Y57A10A  
C:Genetics:  
A:Gene: CESP:Y57A10A.ff  
A:Introns: 34/1; 82/3; 156/3; 398/3; 556/3; 798/1; 863/3; 1024/1

```

Query Match          9.9%  Score 70.5;  DB 2;  Length 1048;
Best Local Similarity 36.7%  Pred. No. 61;
Matches 22;  Conservative 10;  Mismatches 25;  Indels 3;  Gaps 2;

QY  21  TPIHLKGVCALYNEQDEEAAVLEVPQHSDDLHLHCRILIEADP--QTSITLYSMLQLNF 78
      |||:::  :||:::|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db   346  TPKHIECVSVTVRNESDKKHAESSTSHGLSVLYFNKRKMECDPAAKYDITLTKTWVLIQ 404

RESULT  13
T23051
hypothetical protein H0GAl0.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
C:Accession: T23051

```

submitted to the EMBL Data Library, October 1997  
A:Reference number: Z19663  
A:Accession: T23051  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-291 <WIL>  
A:Cross-references: EMBL:Z99773; PIDN:CAB16924.1; GSPDB:GN00028; CESP:H06A10.1  
A:Experimental source: clone H06A10  
C:Genetics:  
A:Gene: CESP:H06A10.1  
A:Map position: X  
A:Introns: 16/1; 59/1; 94/2; 129/3; 189/1; 218/2  
C:Superfamily: Caenorhabditis elegans hypothetical protein H06A10.1

Query Match.	9.7%;	Score 69.5;	DB 2;	Length 291;
Best Local Similarity	23.2%;	Pred. No. 16;		
Matches	29;	Conservative	23;	Mismatches 38; Indels 35; Gaps 6;
QY	14	YFSAGCKTPTHLKDGVCALYNEQDEERAAVLVEVPOHSLSLLHCRIRIADPOTSTLXSM	73	
Db	191	YFFAGAGSEIDQKSG-----KKAAIL-----RSDQTLNYCPMLK-----DITDQDII	232	
QY	74	LQNFENAMRGCCWLADLHNLVRLCFQOSLE--HLDEASFSDIVSGFIEHAAEVRVYA	131	
		:   :   :   :   :   :		
Db	233	IRLDQSOTIYDIIEWISV-----FCYKSHDFGHLD-----MGLVENEQVPPYIP	277	
		:   :   :   :   :   :		
QY	132	QIDES	136	
		:		
Db	278	DISIS	282	

RESULT 14  
S76366  
hypothetical protein - *Synechocystis* sp. (strain PCC 6803)  
C:Species: *Synechocystis* sp.  
A:Variety: PCC 6803  
C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 20-Jun-2000  
C:Accession: S76366

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; *et al.* DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechococcus* sp.

A:Reference number: S74322; MUID:97061201

A:Accession: S76366

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-393 <KAN>

A:Cross-references: EMBL:D64000; GB:AB001339; NID:gl001484; PIDN:BAA10218.1; PID:g...

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C:Superfamily: mannose-1-phosphate guanylyltransferase

	Query Match	9.7%	Score 69.5;	DB 2:	Length 393;
	Best Local Similarity	21.1%;	Pred. No. 23;	Mismatches 62;	Gaps 29;
	Matches 31;	Conservative 25;	Indels 29;	Gaps	
Oy	5	QORVERFIQYFSAGCKTPIHLKDGV-----CALYNQEDEAAVLVPQHSDSL----	53		
Dd	140	QDHLEQLAHFOAKSLSLILLLRRVSNPASFQGVATVNDQGKVLAIVLKPEPPPSNALVG	199		
Oy	54	-----LHCRRIIEADPO-----TSITLYSMILLQNLFENAMRGCVWLDELHNYPCLC	99		
Dd	200	LYFFAPTTHQAIAINLESARGELBITDAIQLISHDYRVESLIQLKGWMLDTGKKDDULLA	259		
Oy	100	FQSLEHLDEASFDIVSGFTEHAEEV	126		
Dd	260	NQIIIDTLVEKN----IQGTVDQSKI	282		

RESULT 15  
ZLNZMV  
genome polyprotein - measles virus (strain Udem)  
N:Contains: RNA-directed RNA polymerase (EC 2.7.7.48)  
C:Species: measles virus  
C:Date: 31-Dec-1989 #sequence\_revision 31-Dec-1989 #t  
C:Accession: A28919  
R:Blumberg, B.M.; Crowley, J.C.; Silverman, J.I.; Men  
Virology 164, 487-497, 1988  
A:Title: Measles virus L protein evidences elements o  
A:Reference number: A28919; MUID:88219537  
A:Accession: A28919  
A:Molecule type: genomic RNA  
A:Residues: 1-2183 <BLU>  
A:Cross-references: GB:M20865; NID:g331776; PIDN:AAA4  
C:Genetics:  
A:Gene: L  
C:Superfamily: parainfluenza virus RNA-directed RNA p  
C:Keywords: ATP; nucleotidyltransferase

	Query Match	9.7%; score 69.5; DB 1;	Length 2183;
	Best Local Similarity	20.0% pred.	No. 1.8e+02;
	Matches 38;	Conservative 26;	Mismatches 63; Indels 63; Gaps 7;
Qy	4	SQQVERELQVFSAGCK-----TPHLKDCVCALYNEQDEEAAVLVP	46
Dd	158	SSSQFEPFLWFVTVKTEMRSVIKSQTHTCHRRRHPTPVFTGSSVELTISRDLVAIIKES	217
Qy	47	QH-----SDSLMLHCRIFHEAD--POTSITLYSMLLQLNFEMAAARGCWGLAD-----	91
Dd	218	QHVVYLTFPELVIMCYDVEIGRLMTETAMTDARYTEL---LGRVRYWKKLIDGFPPALGN	274
Qy	92	-----ELHNVRLE-CFOQSLEHLDCAASDVISVGFIHAEE	125
Dd	275	PTYOIVAMEPLSLAYLOLRDITVELRGAFLNHCFTETHVDLQNGFSD-EGTYHELIE	332
Qy	126	VREYIAOLDE	135
Dd	333	ALDYFITDD	342

us-09-596-784-4.rpr

Wed Jun 6 10:19:17 2001

Search completed: June 5, 2001, 18:18:12  
Job time: 238 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2001, 18:21:11 ; Search time 30.94 seconds  
(without alignments)  
67.946 Million cell updates/sec

Title: US-09-596-784-4

Perfect score: 715

Sequence: 1 MTSSQORVERFLQYFSAGCK.....IEHAAEVREYIAQLDESSAA 139

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 296979 seqs, 15124130 residues

Total number of hits satisfying chosen parameters: 296979

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending\_Patents\_AA\_New:\*  
1: /cgn2.6/ptodata/2/paa/PCT\_NEW\_COMB.pcp.\*  
2: /cgn2.6/ptodata/2/paa/US06\_NEW\_COMB.pcp.\*  
3: /cgn2.6/ptodata/2/paa/US07\_NEW\_COMB.pcp.\*  
4: /cgn2.6/ptodata/2/paa/US08\_NEW\_COMB.pcp.\*  
5: /cgn2.6/ptodata/2/paa/US09\_NEW\_COMB.pcp.\*  
6: /cgn2.6/ptodata/2/paa/US60\_NEW\_COMB.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	68	9.5	652	US-09-813-742-5	Sequence 5, Appli
2	68	9.5	1144	US-09-813-742-3	Sequence 3, Appli
3	66	9.2	466	US-09-739-449-12177	Sequence 12177, A
4	65	9.1	1275	1 PCT-US01-04098A-1583	Sequence 1583, Ap
5	65	9.1	1301	1 PCT-US01-04098A-3551	Sequence 3551, Ap
6	63.5	8.9	1867	5 US-09-824-574-5	Sequence 5, Appli
7	63	8.8	343	5 US-09-270-849B-181413	Sequence 181413,
8	62.5	8.7	947	6 US-09-248-505-702	Sequence 702, App
9	62.5	8.7	968	6 US-09-248-505-1139	Sequence 1139, Ap
10	62	8.7	511	1 PCT-US01-11988-2010	Sequence 2010, Ap
11	62	8.7	511	5 US-09-833-245-2010	Sequence 2010, Ap
12	61.5	8.6	313	5 US-09-270-849B-192462	Sequence 192462,
13	61	8.5	341	5 US-09-739-449-10007	Sequence 10007, A
14	61	8.5	528	6 US-09-248-505-758	Sequence 758, App
15	61	8.5	873	1 PCT-US00-35017A-1214	Sequence 1214, Ap
16	60.5	8.5	1120	1 PCT-US01-04098A-1515	Sequence 1515, Ap
17	60.5	8.5	1152	1 PCT-US01-04098A-3483	Sequence 3483, Ap
18	60	8.4	839	1 PCT-US00-09897-7	Sequence 194981,
19	60	8.4	839	5 US-09-824-258-2	Sequence 2, Appli
20	60	8.4	839	5 US-09-824-258-8	Sequence 8, Appli
21	60	8.4	839	5 US-09-824-258-2	Sequence 8, Appli
22	60	8.4	2263	1 PCT-US01-04098A-1662	Sequence 1662, Ap
23	59.5	8.3	536	1 PCT-US01-04098A-3539	Sequence 3539, Ap
24	59.5	8.3	1047	1 PCT-US01-04098A-1169	Sequence 1169, Ap
25	59.5	8.3	1047	1 PCT-US01-04098A-3137	Sequence 3137, Ap
26	59	8.3	149	5 US-09-270-849B-194537	Sequence 194537,
27	59	8.3	226	5 US-09-270-849B-182172	Sequence 182172,

## ALIGNMENTS

### RESULT 1

US-09-813-742-5  
; Sequence 5, Application US/09813742  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Barbara  
; APPLICANT: Dinesh-Kumar, S.P.  
; TITLE OF INVENTION: NON-PATHOGEN INDUCED SYSTEMIC ACQUIRED RESISTANCE (SAR) IN P  
; FILE REFERENCE: 042250/209601 (5830-12)  
; CURRENT APPLICATION NUMBER: US/09/813,742  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,027  
; PRIOR FILING DATE: 2000-03-21  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 5  
; LENGTH: 652  
; TYPE: PRT  
; ORGANISM: Nicotiana glutinosa  
US-09-813-742-5

Query Match 9.5%; Score 68; DB 5; Length 652;  
Best Local Similarity 23.6%; Pred. No. 39;  
Matches 38; Conservative 23; Mismatches 46; Indels 54; Gaps 8;

Qy	12	LOYFSAGCKTPHLLKGVCAVLEVPQSDSLLHLCRIIEADPQTSITLYS	71
Db	317	LDWFONGSRIITTRDK--HLIEKNDIIEVTALPDHE-----SIOLFK	358
Qy	72	M-----LLOLNFMAAM-----RGCLAL-----DELHNRVLC-FOQSLHLEDAFSFD	114
Db	359	QHAFKQEVNENFEKLSLEVNVYAKGLPLAKVWGSLLHNLRLTENKSAIEHKNNNSYG	418
Qy	115	IVSG-----FIEHA-----AEVREYIAQLDES	136
Db	419	IDKLKISYDGLPEKQEMFLDIACFLRGEKDYILQILES	459

### RESULT 2

US-09-813-742-3  
; Sequence 3, Application US/09813742  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Barbara  
; APPLICANT: Dinesh-Kumar, S.P.  
; TITLE OF INVENTION: NON-PATHOGEN INDUCED SYSTEMIC ACQUIRED RESISTANCE (SAR) IN P  
; FILE REFERENCE: 042250/209601 (5830-12)  
; CURRENT APPLICATION NUMBER: US/09/813,742  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,027  
; PRIOR FILING DATE: 2000-03-21

28	59	8.3	366	5	US-09-270-849B-193802	Sequence 193802,
29	59	8.3	441	5	US-09-270-849B-183122	Sequence 183122,
30	59	8.3	1111	6	US-09-270-849B-183122	Sequence 71, Appl
31	58.5	8.2	107	5	US-09-640-211A-2165	Sequence 2165, Ap
32	58.5	8.2	370	5	US-09-345-236B-133	Sequence 133, App
33	58.5	8.2	557	5	US-09-739-449-12924	Sequence 12924, A
34	58.5	8.2	634	1	PCT-US01-04098A-1066	Sequence 1066, Ap
35	58.5	8.2	656	1	PCT-US01-04098A-3034	Sequence 3034, Ap
36	58.5	8.2	1226	5	US-09-739-449-9308	Sequence 9308, Ap
37	58	8.1	118	5	US-09-270-849B-194271	Sequence 194271,
38	58	8.1	641	5	US-09-768-877-26	Sequence 26, Appl
39	57.5	8.0	111	5	US-09-270-849B-186506	Sequence 186506,
40	57.5	8.0	281	5	US-09-270-849B-189982	Sequence 189982,
41	57.5	8.0	307	5	US-09-826-019-69	Sequence 69, Appl
42	57.5	8.0	615	1	PCT-US01-04098A-3300	Sequence 3300, Ap
43	57.5	8.0	622	1	PCT-US01-04098A-1332	Sequence 1332, Ap
44	57.5	8.0	951	1	PCT-US01-04098A-3596	Sequence 3596, Ap
45	57.5	8.0	1053	1	PCT-US01-04098A-1628	Sequence 1628, Ap

```
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 1144
; TYPE: PRT
; ORGANISM: Nicotiana glauca
; US-09-813-742-3

Query Match          9.5%; Score 68; DB 5; Length 1144;
Best Local Similarity 23.6%; Pred. No. 80;
Matches 38; Conservative 23; Mismatches 46; Indels 54; Gaps 8;

QY 12 LOYFAGCKTPHLKGVCAALYNEQDEAAVLEVPQHSLSLLHCRRIEADPQTSITLYS 71
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 317 LWFNGSGRIITTRDK--HLIEKNDIYEVTPDHE-----SIQLFK 358
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY 72 M-----LLQLNFEMAM-----RCGLAL-----DELHNVRLC-FOOSLEHLEASPSD 114
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 359 QHAFGKEVPNEFEKLSLEVVNAYKGLPLAKVWGSLLHNLRLTEWKSATIEHMKNNYSYG 418
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY 115 IVSG-----FIEHA-----AEVREYIAQDES 136
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 419 IIDKLKISYDGLPKQOEMLDIACFLRGBEKDVILQILES 459
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 3
US-09-739-449-12177
; Sequence 12177, Application US/09739449
; GENERAL INFORMATION:
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15490)C
; CURRENT APPLICATION NUMBER: US/09739,449
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 09/514,000
; PRIOR FILING DATE: 2000-02-23
; NUMBER OF SEQ ID NOS: 13351
; SEQ ID NO 12177
; LENGTH: 466
; TYPE: PRT
; ORGANISM: Agrobacterium tumefaciens
; US-09-739-449-12177

Query Match          9.2%; Score 66; DB 5; Length 466;
Best Local Similarity 25.2%; Pred. No. 42;
Matches 31; Conservative 19; Mismatches 41; Indels 32; Gaps 6;

QY 25 LKDGVCALYNEQDEAAVLEVPQHSLSLLHCRRIEADPQTSITLYSMLLQNFEMAMR 84
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 257 LKTPVAILN-----EARTLE-PQHGDLVRAQADAMAQVQSYLS-----RARIAAOR 303
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY 85 GCWL-----ALDELHNV-----LCFQ-----SLEHLEASFSDIVSGFIEHAE 125
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 304 GSILARVEAEPALERLVRMRLNPKQVLFNFEQPGVILGMEQDLEEVVGNLLENAAR 363
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY 126 VRE 128
DB 364 FAE 366

RESULT 4
PCT-US01-04098A-1583
; Sequence 1583, Application PC/TUS0104098A
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-029
; CURRENT APPLICATION NUMBER: PCT/US01/04098A
; CURRENT FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: Not Yet Assigned
```

```
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 1144
; TYPE: PRT
; ORGANISM: Nicotiana glauca
; US-09-813-742-3

Query Match          9.5%; Score 68; DB 5; Length 1144;
Best Local Similarity 23.6%; Pred. No. 80;
Matches 38; Conservative 23; Mismatches 46; Indels 54; Gaps 8;

QY 12 LOYFAGCKTPHLKGVCAALYNEQDEAAVLEVPQHSLSLLHCRRIEADPQTSITLYS 71
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 317 LWFNGSGRIITTRDK--HLIEKNDIYEVTPDHE-----SIQLFK 358
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY 72 M-----LLQLNFEMAM-----RCGLAL-----DELHNVRLC-FOOSLEHLEASPSD 114
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 359 QHAFGKEVPNEFEKLSLEVVNAYKGLPLAKVWGSLLHNLRLTEWKSATIEHMKNNYSYG 418
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY 115 IVSG-----FIEHA-----AEVREYIAQDES 136
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 419 IIDKLKISYDGLPKQOEMLDIACFLRGBEKDVILQILES 459
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 3
US-09-739-449-12177
; Sequence 12177, Application US/09739449
; GENERAL INFORMATION:
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15490)C
; CURRENT APPLICATION NUMBER: US/09739,449
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 09/514,000
; PRIOR FILING DATE: 2000-02-23
; NUMBER OF SEQ ID NOS: 13351
; SEQ ID NO 12177
; LENGTH: 466
; TYPE: PRT
; ORGANISM: Agrobacterium tumefaciens
; US-09-739-449-12177

Query Match          9.2%; Score 66; DB 5; Length 466;
Best Local Similarity 25.2%; Pred. No. 42;
Matches 31; Conservative 19; Mismatches 41; Indels 32; Gaps 6;

QY 25 LKDGVCALYNEQDEAAVLEVPQHSLSLLHCRRIEADPQTSITLYSMLLQNFEMAMR 84
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 257 LKTPVAILN-----EARTLE-PQHGDLVRAQADAMAQVQSYLS-----RARIAAOR 303
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY 85 GCWL-----ALDELHNV-----LCFQ-----SLEHLEASFSDIVSGFIEHAE 125
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 304 GSILARVEAEPALERLVRMRLNPKQVLFNFEQPGVILGMEQDLEEVVGNLLENAAR 363
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY 126 VRE 128
DB 364 FAE 366

RESULT 4
PCT-US01-04098A-1583
; Sequence 1583, Application PC/TUS0104098A
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-029
; CURRENT APPLICATION NUMBER: PCT/US01/04098A
; CURRENT FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: Not Yet Assigned
```

```
; NAME/KEY: misc_feature
; LOCATION: (1)...(1301)
; OTHER INFORMATION: xaa = any amino acid or nothing
PCT-US01-04098A-3551

Query Match      9.1%; Score 65; DB 1; Length 1301;
Best Local Similarity 25.9%; Pred. No. 2e+02;
Matches 30; Conservative 28; Mismatches 46; Indels 12; Gaps 6;

Qy 30 CALYNEQDEAAVLEVPQHSLSLL-HC-----RIIEADPQTSITLYSMLQLNPFEMAAAMR 84
Db 607 CERTQOHHEAMKTIQI---RESLLAKHALEKQOLFAYERTHQLRSELDKLNKEVTAVQ 663
Qy 85 GCLWAL-DELHNVRCLFQOSLEHLDSEAFSDIVSGFIEHAAEVREYIAQLDESSAA 139
Db 664 ECVLEVCREKNDLELTLRKTEK-EQOTQEKIKELIQOLE--KEWQSKLDQTIKA 716

RESULT 6
US-09-824-574-5
; Sequence 5, Application US/09824574
; GENERAL INFORMATION:
; APPLICANT: Rouleau, Natalie
; APPLICANT: Moilanen, Anu-Maarit
; APPLICANT: Palvimäki, Jorma J.
; APPLICANT: Jänne, Olli A.
; TITLE OF INVENTION: ARIP4 Gene and Protein
; FILE REFERENCE: 2630-109
; CURRENT APPLICATION NUMBER: US/09/824,574
; CURRENT FILING DATE: 2001-04-03
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 1867
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-824-574-5

Query Match      8.9%; Score 63.5; DB 5; Length 1867;
Best Local Similarity 25.2%; Pred. No. 4.7e+02;
Matches 37; Conservative 27; Mismatches 50; Indels 33; Gaps 9;

Qy 12 LOYFAGCKTPHILKGV-----CALY--NEODEE---AAVLEVPQHSLSLLHCRII 59
Db 430 IRYF-VSIKTNFLFAHGLLENVVRIVLYGLNQSDDDVQSVAAASLLTPTSEFVKLNSTI 488
Qy 60 EADPQTSITLYSMLQLNPFEMAAAMRGCLALDELHNVRCLFQ-----SLEHLDEA 110
Db 489 EI---LVTTIWSLLARLDODDSSVGSIMDL---LAKLCHQEBVLDLKKNKALEHPSEW 541
Qy 111 SFSDIVS---GFIEHA-AEVREYIAQL 133
Db 542 SFKSLPKLYPFLRHSTSSVRRVAVLNL 568

RESULT 7
US-09-270-849B-181413
; Sequence 18113, Application US/09270849B
; GENERAL INFORMATION:
; APPLICANT: Swimmer et al.
; TITLE OF INVENTION: Insect genome survey devices
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/09/270,849B
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 195450
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 181413
; LENGTH: 343
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-270-849B-181413

Query Match      8.8%; Score 63; DB 5; Length 343;
Best Local Similarity 18.4%; Pred. No. 60;
Matches 26; Conservative 27; Mismatches 52; Indels 36; Gaps 4;

Qy 4 SQQRVERFLQYFAGCKTPHILKGVCAALYNEQDEAAVLEVPQHSLSLLHCRII---- 59
Db 59 SEQLRRAL-----KTVL-----VYAESDSLQDTSFPEQVQDILLFNJHMLISDT 103
Qy 60 -----EADPQTSITLYSMLQLNPFEMAAAMRGCLALDELHNVRCLFQOSLEHLDSEAFS 113
Db 104 VKMEYQEDPEMLLDLMNRIAGYONNPDRLTWLE-----NNAKKHRRANHT 152
Qy 114 DIVSGFIEHAAEVREYIAQLD 134
Db 153 EAAMCYVHAASLVSEYLSMLE 173

RESULT 8
US-60-248-505-702
; Sequence 702, Application US/60248505
; GENERAL INFORMATION:
; APPLICANT: Beasley, Ellen
; TITLE OF INVENTION: ISOLATED HUMAN G-PROTEIN COUPLED
; TITLE OF INVENTION: RECEPTORS, NUCLEIC ACID MOLECULES ENCODING HUMAN GPCR
; FILE REFERENCE: c1000918
; CURRENT APPLICATION NUMBER: US/60/248,505
; CURRENT FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 1998
; SOFTWARE: FastSeq for Windows version 4.0
; SEQ ID NO 702
; LENGTH: 947
; TYPE: PRT
; ORGANISM: Human
US-60-248-505-702

Query Match      8.7%; Score 62.5; DB 6; Length 947;
Best Local Similarity 25.8%; Pred. No. 2.5e+02;
Matches 34; Conservative 21; Mismatches 48; Indels 29; Gaps 7;

Qy 25 LKDGVCALYNEQDEAAVLEVP--QHSLSLLHC-RIIEADPQTSITLYSMLQLNPFEMA 81
Db 290 LKNDITAVTKEGKILLTNLEVPDTEGAVSSRLCHQISGDWOTINKLLTQVHDMADF-- 347
Qy 82 AMRGCLALDELHNVRCL-----CFQOSLEHL-----DEAFSDIVSGFIEHAEE 125
Db 348 ---GFN---EKHQLKMEQYLQWLKFEQDFQVLTVEFEVLLNQQAELAD-VTGTLAQVKQ 399
Qy 126 VREYIAQLDESS 137
Db 400 KIKKLENDENS 411

RESULT 9
US-60-248-505-1139
; Sequence 1139, Application US/60248505
; GENERAL INFORMATION:
; APPLICANT: Beasley, Ellen
; TITLE OF INVENTION: ISOLATED HUMAN G-PROTEIN COUPLED
; TITLE OF INVENTION: RECEPTORS, NUCLEIC ACID MOLECULES ENCODING HUMAN GPCR
; FILE REFERENCE: c1000918
; CURRENT APPLICATION NUMBER: US/60/248,505
; CURRENT FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 1998
; SOFTWARE: FastSeq for Windows version 4.0
; SEQ ID NO 1139
; LENGTH: 968
```



```

Db      160 P-----VNLGIAQRMNLNHEKEAWNNIIVRHQK-SFRDKFRYDIDKIYRDLQ 206
Qy      111 SFSDIVSGFTEHAAE 125
        | : : || : |
Db      207 KFENLTTKFIKIAEE 221

```

```

RESULT 13
US-09-739-449-10007
; Sequence 10007, Application US/09739449
; GENERAL INFORMATION:
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(115490)C
; CURRENT APPLICATION NUMBER: US/09/739,449
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 09/514,000
; PRIOR FILING DATE: 2000-02-23
; NUMBER OF SEQ ID NOS: 13351
; SEQ ID NO 10007
; LENGTH: 341
; TYPE: PRT
; ORGANISM: Agrobacterium tumefaciens
US-09-739-449-10007

```

Query Match	8.5%	Score 61	DB 5	Length 341
Best Local Similarity	23.1%	Pred. No. 98		
Matches	27	Conservative	21	Mismatches 43
				Indels 26
				Gaps 6
Qy	25	LKDGVCALY-----NEQDEEAALVEVPOHSDSLLLHCRIRIADPOTSTITLSMLLQL	76	
		:    :	:	:
		:	:	:
Db	38	LGGGICATIPSGDLVLQEPDEQ-----HYDTPL--AAPLERQPSLDIAMEKAAL	88	
		:	:	:
		:	:	:
Qy	77	NFEMAARGCWIALDEHNVRLCTQQ-----SLEHLD--EAFSDIVSGFISHAAEVR	127	
		:	:	:
		:	:	:
Db	89	RLDRA---GAWRLCRSTADIKKAFAEDIFAALVAMEGCGEATGADLDALEVYFAAGLR	142	
		:	:	:
		:	:	:

```

RESULT 14
US-60-248-505-758
; Sequence 758, Application US/60248505
; GENERAL INFORMATION:
; APPLICANT: Beasley, Ellen
; TITLE OF INVENTION: ISOLATED HUMAN G-PROTEIN COUPLED
; TITLE OF INVENTION: RECEPTORS, NUCLEIC ACID MOLECULES
; TITLE OF INVENTION: PROTEINS, AND USES THEREOF
; FILE REFERENCE: gl000918
; CURRENT APPLICATION NUMBER: US/60/248,505
; CURRENT FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 1998
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 758
; LENGTH: 528
; TYPE: PRT
; ORGANISM: Human
US-60-248-505-758

```

	Query Match	8.5%	Score 61;	DB 6;	Length 528;
	Best Local Similarity	27.7%;	pred. No.	1.7e+02;	
	Matches	Conservative	17;	Mismatches	42; Indels 40; Gaps 9;
Qy	22	PIHKDGVCALYNQDEFAAVLEV-----POHSDLLHCRIILRADPOTSITL	69		
		( :   :   )			
Dd	108	PVH-EVGECARAIE-DKDNVEYLESHMLSNPASPYENCFNLCJCLELEG-AQELVT-	163		
Qy	70	YSMLQLNFEMAARGCMALDELH-----NVRLCFQOSLEHLUDEASFSDIVSGFTEH	122		
		( :   :   )			
Dd	164	-----QKAMGVSAGDLLHQLOSQVTNASLT-LKLADSQCSE-----GALOQ	208		
Qy	123	AAEVREYIAQLDESAA	139		

```

Db      209 EVDVLE--SQLSESSCA 223
      :| | :| | | | |

```

```

RESULT 15
PCT-US00-35017A-1214
; Sequence 1214, Application PC/TUS0035017A
; GENERAL INFORMATION:
; APPLICANT: Hyseq Inc
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 784PCT
; CURRENT APPLICATION NUMBER: PCT/US00/35017A
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US09/552,317
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 1478
; SEQ ID NO 1214
; LENGTH: 873
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US00-35017A-1214

```

Query Match	8.5%;	Score 61;	DB 1;	Length 873;
Best Local Similarity	23.0%;	Pred. No. 3.3e+02;		
Matches	29;	Conservative 19;	Mismatches 56;	Indels 22; .Caps
Qy	20	KTPHLLKDGVCALYNEQDEEAA--VLEVPQHSDDLHLHCRILIEADPOTSITLYSMILLQLN	77	
Db	513	KLVVHLYH-LQMPFSPESADIAKFLVRDAMHEEMETETKGRALPGDLVLIVLYKLTGLL	571	
Qy	78	FEMAAMRGCMALDELHNV---RLCFOQSLEHLDEASFSDIVSG-----FIEHAAEVR	127	
Db	572	FPTS-----DFWHPVVTPALVCLSQLLTCKPILSLQDVYKGLFVCCLCFLEYVALSQ	622	
Qy	128	EYIAQL	133	
Db	623	RFIPEL	628	

Search completed: June 5, 2001, 18:21:12  
Job time: 388 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2001, 18:20:30 ; Search time 139.63 seconds  
(without alignments)  
160.129 Million cell updates/sec

Title: US-09-596-784-4

Perfect score: 715

Sequence: 1 MTSSQQRVERFLQYFSAGCK.....IEHAAEVREYIAQLDESSAA 139

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1009251 seqs, 160854530 residues

Total number of hits satisfying chosen parameters: 1009251

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending\_Patents\_AA\_Main:\*

- 1: /cgn2\_6/ptodata/2/paa/US06\_COMB.pcp.\*
- 2: /cgn2\_6/ptodata/2/paa/US06\_COMB.pcp.\*
- 3: /cgn2\_6/ptodata/2/paa/US07\_COMB.pcp.\*
- 4: /cgn2\_6/ptodata/2/paa/US08\_COMB.pcp.\*
- 5: /cgn2\_6/ptodata/2/paa/US081\_COMB.pcp.\*
- 6: /cgn2\_6/ptodata/2/paa/US082\_COMB.pcp.\*
- 7: /cgn2\_6/ptodata/2/paa/US083\_COMB.pcp.\*
- 8: /cgn2\_6/ptodata/2/paa/US084\_COMB.pcp.\*
- 9: /cgn2\_6/ptodata/2/paa/US085\_COMB.pcp.\*
- 10: /cgn2\_6/ptodata/2/paa/US086\_COMB.pcp.\*
- 11: /cgn2\_6/ptodata/2/paa/US087\_COMB.pcp.\*
- 12: /cgn2\_6/ptodata/2/paa/US088\_COMB.pcp.\*
- 13: /cgn2\_6/ptodata/2/paa/US089\_COMB.pcp.\*
- 14: /cgn2\_6/ptodata/2/paa/US090\_COMB.pcp.\*
- 15: /cgn2\_6/ptodata/2/paa/US091\_COMB.pcp.\*
- 16: /cgn2\_6/ptodata/2/paa/US092\_COMB.pcp.\*
- 17: /cgn2\_6/ptodata/2/paa/US093\_COMB.pcp.\*
- 18: /cgn2\_6/ptodata/2/paa/US094\_COMB.pcp.\*
- 19: /cgn2\_6/ptodata/2/paa/US095\_COMB.pcp.\*
- 20: /cgn2\_6/ptodata/2/paa/US096\_COMB.pcp.\*
- 21: /cgn2\_6/ptodata/2/paa/US097\_COMB.pcp.\*
- 22: /cgn2\_6/ptodata/2/paa/US098\_COMB.pcp.\*
- 23: /cgn2\_6/ptodata/2/paa/US099\_COMB.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	715	100.0	139	1	PCT-US98-15426-4
2	715	100.0	139	15	US-09-120-663-4
3	715	100.0	139	18	US-09-412-100-30
4	715	100.0	139	18	US-09-431-614-10
5	715	100.0	139	19	US-09-596-784-4
6	85.5	12.0	1007	3	US-07-861-3900-7
7	76.5	10.7	194	9	US-08-583-808-4
8	76.5	10.7	1105	1	PCT-US96-14679-2
9	76.5	10.7	1105	1	PCT-US96-14679A-2
10	76.5	10.7	1105	16	US-09-220-157-2

11	76.5	10.7	1105	16	US-09-220-157A-2
12	75	10.5	243	1	PCT-US97-07950-338
13	75	10.5	243	12	US-08-858-207A-338
14	75	10.5	317	1	PCT-US00-01132-2
15	75	10.5	317	1	PCT-US00-17262-50
16	75	10.5	317	1	PCT-US97-14436-492
17	75	10.5	317	13	US-08-911-503-492
18	75	10.5	317	13	US-08-911-503A-492
19	75	10.5	317	16	US-09-238-477-2
20	75	10.5	317	19	US-09-595-940-50
21	75	10.5	2053	23	US-60-191-637-39778
22	75	10.5	2053	23	US-60-191-681-30805
23	75	10.5	2053	19	US-09-540-236-3398
24	74	10.3	589	23	US-60-128-476-3291
25	73.5	10.3	181	23	US-60-170-374-3340
26	73.5	10.3	211	23	US-60-178-307-2751
27	73.5	10.3	984	8	US-08-449-645-19
28	73.5	10.3	984	11	US-08-702-367-19
29	73.5	10.3	984	17	US-09-378-759-19
30	73	10.2	217	18	US-09-489-039A-7789
31	73	10.2	560	18	US-09-450-969-4699
32	73	10.2	662	23	US-60-207-422-133
33	73	10.2	674	23	US-60-212-413-305
34	73	10.2	674	23	US-60-229-518-252
35	72.5	10.1	244	21	US-09-733-089-23131
36	71.5	10.0	425	16	US-09-248-796-14511
37	71.5	10.0	920	23	US-60-222-113-76
38	70.5	9.9	103	15	US-09-107-532-5377
39	70.5	9.9	1445	21	US-09-733-089-22848
40	70.5	9.9	2183	16	US-09-269-367A-4
41	70.5	9.9	2183	16	US-09-269-367A-12
42	69.5	9.7	286	12	US-08-827-356-3003
43	69.5	9.7	286	20	US-09-611-529-3871
44	69.5	9.7	2183	16	US-09-269-367A-2
45	69.5	9.7	2183	16	US-09-269-367A-6

## ALIGNMENTS

RESULT 1

PCT-US98-15426-4

; Sequence 4, Application PC/TUS9815426

; GENERAL INFORMATION:

; APPLICANT: Cornell Research Foundation, Inc.

; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR FROM

; TITLE OF INVENTION: ERWINIA AMYLOVORA, ITS USE, AND ENCODING GENE

; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP

; STREET: P.O. Box 1051, Clinton Square

; CITY: Rochester

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 14603

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US98/15426

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/055,105

; FILING DATE: 06-AUG-1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Goldman, Michael L.

; REGISTRATION NUMBER: 30,727

; REFERENCE/DOCKET NUMBER: 19603/1662

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (716) 263-1304

```
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 139 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US98-15426-4

Query Match      100.0%; Score 715; DB 1; Length 139;
Best Local Similarity 100.0%; Pred. No. 2.8e-75;
Matches 139; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTSSQQRVERFLQYFSAGCKTPHILKDGVCALYNEQDEAAVLEVPQHSDDLHCRRIE 60
DB 1 MTSSQQRVERFLQYFSAGCKTPHILKDGVCALYNEQDEAAVLEVPQHSDDLHCRRIE 60
QY 61 ADPQTSITLYSMLLQNFEMAAAMRGCVLALDELHNRLCFQOSLEHLDEASFSDIVSGFI 120
DB 61 ADPQTSITLYSMLLQNFEMAAAMRGCVLALDELHNRLCFQOSLEHLDEASFSDIVSGFI 120
QY 121 EHAEEVREYIAQLDESSAA 139
DB 121 EHAEEVREYIAQLDESSAA 139

RESULT 2
US-09-120-663-4
; Sequence 4, Application US/09120663
; GENERAL INFORMATION:
; APPLICANT: Bogdanove, Adam J.
; APPLICANT: Kim, Jihyun Francis
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Beer, Steven V.
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR FROM
; TITLE OF INVENTION: ERWINIA AMYLOVORA, ITS USE, AND ENCODING GENE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: P.O. Box 1051, Clinton Square
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/120,663
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/055,105
; FILING DATE: 06-AUG-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/1661
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 139 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-120-663-4
```

```
Query Match      100.0%; Score 715; DB 15; Length 139;
Best Local Similarity 100.0%; Pred. No. 2.8e-75;
Matches 139; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTSSQQRVERFLQYFSAGCKTPHILKDGVCALYNEQDEAAVLEVPQHSDDLHCRRIE 60
DB 1 MTSSQQRVERFLQYFSAGCKTPHILKDGVCALYNEQDEAAVLEVPQHSDDLHCRRIE 60
QY 61 ADPQTSITLYSMLLQNFEMAAAMRGCVLALDELHNRLCFQOSLEHLDEASFSDIVSGFI 120
DB 61 ADPQTSITLYSMLLQNFEMAAAMRGCVLALDELHNRLCFQOSLEHLDEASFSDIVSGFI 120
QY 121 EHAEEVREYIAQLDESSAA 139
DB 121 EHAEEVREYIAQLDESSAA 139

RESULT 3
US-09-412-100-30
; Sequence 30, Application US/09412100
; GENERAL INFORMATION:
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Fan, Hao
; APPLICANT: Niggemeyer, Jennifer L.
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR FRAGMENTS WHICH ARE
; TITLE OF INVENTION: ACTIVE BUT DO NOT ELICIT A HYPERSENSITIVE RESPONSE
; FILE REFERENCE: 21829/31 (EBC-002)
; CURRENT APPLICATION NUMBER: US/09/412,100
; CURRENT FILING DATE: 1999-10-04
; EARLIER APPLICATION NUMBER: 60/103,050
; EARLIER FILING DATE: 1998-10-05
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 30
; LENGTH: 139
; TYPE: PRT
; ORGANISM: Erwinia amylovora
; US-09-412-100-30

Query Match      100.0%; Score 715; DB 18; Length 139;
Best Local Similarity 100.0%; Pred. No. 2.8e-75;
Matches 139; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTSSQQRVERFLQYFSAGCKTPHILKDGVCALYNEQDEAAVLEVPQHSDDLHCRRIE 60
DB 1 MTSSQQRVERFLQYFSAGCKTPHILKDGVCALYNEQDEAAVLEVPQHSDDLHCRRIE 60
QY 61 ADPQTSITLYSMLLQNFEMAAAMRGCVLALDELHNRLCFQOSLEHLDEASFSDIVSGFI 120
DB 61 ADPQTSITLYSMLLQNFEMAAAMRGCVLALDELHNRLCFQOSLEHLDEASFSDIVSGFI 120
QY 121 EHAEEVREYIAQLDESSAA 139
DB 121 EHAEEVREYIAQLDESSAA 139

RESULT 4
US-09-431-614-10
; Sequence 10, Application US/09431614
; GENERAL INFORMATION:
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Schading, Richard L.
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR-INDUCED STRESS
; TITLE OF INVENTION: RESISTANCE
; FILE REFERENCE: 21829/41 (EBC-003)
; CURRENT APPLICATION NUMBER: US/09/431,614
; CURRENT FILING DATE: 1999-11-02
; EARLIER APPLICATION NUMBER: 60/107,243
; EARLIER FILING DATE: 1998-11-05
; NUMBER OF SEQ ID NOS: 18
```

Query Match 100.0%; Score 715; DB 19; Length 139;

Qy 109 E-ASFSDIVSG 118  
Db 215 GLAOFPPDILPG 225

RESULT 8  
 PECT-US96-14679-2  
 : Sequence 2, Application PC/TUS9614679  
 : GENERAL INFORMATION:  
 : APPLICANT: VILLEPONTEAU, BRYANT

RESULT 9  
PCT-US96-14679A-2  
; Sequence 2, Application PC/TUS9614679A  
; GENERAL INFORMATION:  
; APPLICANT: VILLEPONTEAU, BRYANT  
; APPLICANT: FENG, JUNLI  
; APPLICANT: ANDREWS, WILLIAM H.  
; APPLICANT: ADAMS, ROBERT R.  
; TITLE OF INVENTION: METHODS AND REAGENTS FOR REGULATING  
; TITLE OF INVENTION: TELOMERE LENGTH AND TELOMERASE ACTIVITY  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GERON CORPORATION  
; STREET: 200 CONSTITUTION DRIVE  
; CITY: MENLO PARK  
; STATE: CALIFORNIA  
; COUNTRY: U.S.A.  
; ZIP: 94025

```

RESULT 11
US-09-220-157A-2
; Sequence 2, Application US/09220157A
; GENERAL INFORMATION:
; APPLICANT: Villeponteau, Bryant
; APPLICANT: Feng, Junli
; APPLICANT: Adams, William H.
; APPLICANT: Adams, Robert R.
; TITLE OF INVENTION: Methods and Reagents for Regulating
; TITLE OF INVENTION: Telomere Length and Telomerase Activity
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/220.157A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/710,249
; FILING DATE: 13-SEP-1996

```





TELEX:  
; INFORMATION FOR SEQ ID NO: 338:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 243 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: None  
US-08-858-207A-338

Query Match 10.5%; Score 75; DB 12; Length 243;  
Best Local Similarity 23.8%; Pred. No. 2.6;  
Matches 29; Conservative 25; Mismatches 30; Indels 38; Gaps 7;  
  
Qy 20 KTPHLKDG--VCA-----LYNEQDEEAAVLEVPQHSDSLLHLC 56  
Db 106 KKPISRDGMKLCVETSTTFDDWVRQSEKDYODMLIYLKENDFAKIGELTE-KNALAMH- 163  
  
Qy 57 RIIEADPOTSITLYSMILQLNFE-MAAMR-----GWLALDELHNRL-CFOQSLEHL 107  
Db 164 ----ATTKTASPAFSYLTDSAYEAMAFVRLREKGEACYFTMDAGPNVKVFCOEKDLHL 219  
  
Qy 108 DE 109  
Db 220 SE 221

RESULT 14  
PCT-US00-01132-2  
; Sequence 2, Application PC/TUS0001132  
; GENERAL INFORMATION:  
; APPLICANT: SmithKline Beecham Corporation  
; TITLE OF INVENTION: mvd  
; FILE REFERENCE: GM10188  
; CURRENT APPLICATION NUMBER: PCT/US00/01132  
; CURRENT FILING DATE: 2000-01-19  
; EARLIER APPLICATION NUMBER: US 09/238,477  
; EARLIER FILING DATE: 1999-01-28  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 317  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
PCT-US00-01132-2

Query Match 10.5%; Score 75; DB 1; Length 317;  
Best Local Similarity 23.8%; Pred. No. 3.8;  
Matches 29; Conservative 25; Mismatches 30; Indels 38; Gaps 7;  
  
Qy 20 KTPHLKDG--VCA-----LYNEQDEEAAVLEVPQHSDSLLHLC 56  
Db 180 KKPISRDGMKLCVETSTTFDDWVRQSEKDYODMLIYLKENDFAKIGELTE-KNALAMH- 237  
  
Qy 57 RIIEADPOTSITLYSMILQLNFE-MAAMR-----GWLALDELHNRL-CFOQSLEHL 107  
Db 238 ----ATTKTASPAFSYLTDSAYEAMAFVRLREKGEACYFTMDAGPNVKVFCOEKDLHL 293  
  
Qy 108 DE 109  
Db 294 SE 295

RESULT 15  
PCT-US00-17262-50  
; Sequence 50, Application PC/TUS0017262  
; GENERAL INFORMATION:  
; APPLICANT: SMITHKLINE BEECHAM COPORATION  
; APPLICANT: SMITHKLINE BEECHAM P.I.C.  
; TITLE OF INVENTION: MEVALONATE PATHWAY GENES  
; FILE REFERENCE: GM50062

; CURRENT APPLICATION NUMBER: PCT/US00/17262  
; CURRENT FILING DATE: 2000-06-22  
; PRIOR APPLICATION NUMBER: US 60/140,519  
; PRIOR FILING DATE: 1999-06-22  
; PRIOR APPLICATION NUMBER: US 60/146,682  
; PRIOR FILING DATE: 1999-08-02  
; NUMBER OF SEQ ID NOS: 72  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 50  
; LENGTH: 317  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
PCT-US00-17262-50

Query Match 10.5%; Score 75; DB 1; Length 317;  
Best Local Similarity 23.8%; Pred. No. 3.8;  
Matches 29; Conservative 25; Mismatches 30; Indels 38; Gaps 7;  
  
Qy 20 KTPHLKDG--VCA-----LYNEQDEEAAVLEVPQHSDSLLHLC 56  
Db 180 KKPISRDGMKLCVETSTTFDDWVRQSEKDYODMLIYLKENDFAKIGELTE-KNALAMH- 237  
  
Qy 57 RIIEADPOTSITLYSMILQLNFE-MAAMR-----GWLALDELHNRL-CFOQSLEHL 107  
Db 238 ----ATTKTASPAFSYLTDSAYEAMAFVRLREKGEACYFTMDAGPNVKVFCOEKDLHL 293  
  
Qy 108 DE 109  
Db 294 SE 295

Search completed: June 5, 2001, 18:20:31  
Job time: 357 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2001, 18:16:59 ; Search time 26.26 Seconds  
(without alignments)  
101.687 Million cell updates/sec

Title: US-09-596-784-4

Perfect score: 715

Sequence: 1 MTSSQORVERFLOYSAGCK.....IEHAAEVREYIAQLDESSAA 139

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 185757 seqs, 19210857 residues

Total number of hits satisfying chosen parameters: 185757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	80.5	11.3	368	2	US-08-869-137-2
2	76.5	10.7	1105	2	US-08-710-249-2
3	73.5	10.3	984	2	US-08-673-789-9
4	73.5	10.3	984	2	US-08-449-645A-19
5	73.5	10.3	984	2	US-08-702-367A-19
6	73.5	10.3	984	5	PCT-US95-04681-19
7	69.5	9.7	2183	1	US-08-348-891A-7
8	69.5	9.7	2183	2	US-08-905-817-7
9	68.5	9.6	591	1	US-08-179-738-5
10	68.5	9.6	591	1	US-08-179-738-10
11	68.5	9.6	591	2	US-08-628-145-5
12	68.5	9.6	591	2	US-08-171-718-16
13	68.5	9.6	595	3	US-08-478-087-16
14	68.5	9.6	596	1	US-08-179-738-2
15	68.5	9.6	596	1	US-08-179-738-3
16	68.5	9.6	596	2	US-08-628-145-2
17	68.5	9.6	596	2	US-08-628-145-3
18	68.5	9.6	596	2	US-08-628-145-3
19	68	9.5	652	1	PCT-US95-07754A-6
20	68	9.5	652	5	PCT-US95-07754A-6
21	68	9.5	1143	2	US-08-310-912A-108
22	68	9.5	1143	5	PCT-US95-04589-108
23	68	9.5	1144	1	US-08-261-663A-2
24	68	9.5	1144	1	US-08-261-663A-4
25	68	9.5	1144	3	US-08-930-996A-9
26	68	9.5	1144	5	PCT-US95-07754A-2
27	68	9.5	1144	5	PCT-US95-07754A-4

28 67.5 9.4 584 1 US-08-179-738-7  
29 67.5 9.4 584 2 US-08-628-145-7  
30 65.5 9.2 947 2 US-08-500-857A-4  
31 64.5 9.0 1151 1 US-08-286-889-37  
32 64.5 9.0 1151 1 US-08-485-618-37  
33 64.5 9.0 1151 1 US-08-362-652-37  
34 64.5 9.0 1151 2 US-08-605-672-37  
35 64.5 9.0 1151 2 US-08-482-293A-37  
36 64.5 9.0 1151 2 US-08-943-363-37  
37 64.5 9.0 1161 1 US-08-485-618-55  
38 64.5 9.0 1161 1 US-08-362-652-37  
39 64.5 9.0 1161 2 US-08-605-672-55  
40 64.5 9.0 1161 2 US-08-482-293A-55  
41 64.5 9.0 1161 2 US-08-943-363-55  
42 63.5 8.9 626 1 US-07-938-782A-2  
43 63.5 8.9 626 5 US-08-630-524-2  
44 63.5 8.9 626 5 PCT-US93-08131-2  
45 62.5 8.7 615 2 US-08-484-101B-38

## ALIGNMENTS

RESULT 1  
US-08-869-137-2  
; Sequence 2, Application US/08869137  
; Patent No. 5856157  
; GENERAL INFORMATION:  
; APPLICANT: Schultz, David  
; APPLICANT: Craig, Richard  
; APPLICANT: Medford, June I.  
; APPLICANT: Mumma, R.O.  
; APPLICANT: Cox-Foster, Diana L.  
; TITLE OF INVENTION: A No. 5856157el D9 14:0-ACP Fatty  
; TITLE OF INVENTION: Acid Desaturase and Gene Therefor  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: The Webb Law Firm  
; STREET: 700 Koppers Building, 436 Seventh Avenue  
; CITY: Pittsburgh  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 15219-1818  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/869,137  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/018,957  
; FILING DATE: 04-JUN-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Johnson, Barbara E  
; REGISTRATION NUMBER: 31,198  
; REFERENCE/DOCKET NUMBER: 2034-970766  
; TELEPHONE: 412-471-8815  
; TELEFAX: 412-471-4094  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 368 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FRAGMENT TYPE: internal  
US-08-869-137-2

Sequence 7, Appli  
Sequence 7, Appli  
Sequence 4, Appli  
Sequence 37, Appli  
Sequence 37, Appli  
Sequence 37, Appli  
Sequence 37, Appli  
Sequence 37, Appli  
Sequence 37, Appli  
Sequence 37, Appli  
Sequence 55, Appli  
Sequence 55, Appli  
Sequence 55, Appli  
Sequence 55, Appli  
Sequence 55, Appli  
Sequence 2, Appli  
Sequence 2, Appli  
Sequence 2, Appli  
Sequence 38, Appli

us-09-596-784-4.ra1

Wed Jun 6 10:19:15 2001

OTHER INFORMATION: /note= "deduced amino acid sequence of  
; OTHER INFORMATION: TPC2 open reading frame"  
US-08-710-249-2

8;

Query Match 11.3%; Score 80.5; DB 2; Length 368;  
Best Local Similarity 20.1%; Pred. No. 0.066; 54; Indels 67; Gaps  
Matches 37; Conservative 26; Mismatches 54; Indels 67; Gaps  
QY 6 QVRFYFQYFSA-----GC-KTPIHLKDGVCALYNQDEEAAVLE-----VPQHSDSL 52  
DB 169 ROVETIOYLIAGQDIGTEKNPYHL-----FYTTSFQERATFISHANTAKLAQOQHGDKQ 223  
QY 53 LLH-CRIIEADPQTSITLYSMQLNFEM-----80  
DB 224 LAQICGTIAADEKRHEATVTRIVDKLFDLPDETMSCLAHMMKRKITMPAHLMRDGRDPH 283  
QY 81 -----NAMRCGCWALDELHNVRCLFQOSLEH-LDEASFSDIVSGFIEHAAEVREYIA 131  
DB 284 LFOHFSVASRTGVTYVMDYIN-----ILEHFVEKNIEKITAGLSKDGKREAOQDYVC 335  
QY 132 OLDE 135  
DB 336 KLGE 339

RESULT 2

US-08-710-249-2  
; Sequence 2, Application US/08710249  
; Patent No. 5858777  
; GENERAL INFORMATION:  
; APPLICANT: Villeponteau, Bryant  
; APPLICANT: Feng, Junli  
; APPLICANT: Adams, Robert R.  
; TITLE OF INVENTION: Methods and Reagents for Regulating  
; TITLE OF INVENTION: Telomere Length and Telomerase Activity  
; NUMBER OF SEQUENCES: 26  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08710,249  
; FILING DATE: 13-SEP-1996  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/593,808  
; FILING DATE: 05-JAN-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/003,492  
; FILING DATE: 08-SEP-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Scorella, John R.  
; REGISTRATION NUMBER: 32,944  
; REFERENCE/DOCKET NUMBER: 015389-001220US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 576-0200  
; TELEFAX: (415) 576-0300  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1105 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FEATURE:  
; NAME/KEY: Protein  
; LOCATION: 1..1105

Query Match 10.7%; Score 76.5; DB 2; Length 1105;  
Best Local Similarity 20.0%; Pred. No. 1;  
Matches 25; Conservative 31; Mismatches 46; Indels 23; Gaps 3;  
QY 25 LKDGVCALYNQDEEAAVLEV-----PQHSDSLHCRITIEADPQTSITLYSML 73  
DB 691 LEDKIRALKENKQDQLESLVLELHMQEQYRDPQPHLEKTAQOQKQLED-----LVHIR 744  
QY 74 LQNFEMAAMRCGCWALDELHNVRCLFQOSLEHLDSEAFSDIVSGFIEHAAEVREYIAQL 133  
DB 745 AELSRESTEMENAWNEYLKLENDVEQLKQTLQOHHRAF-----FFQEKSOIQKDLWRI 798  
QY 134 DESSA 138  
DB 799 EDVTA 803

RESULT 3

US-08-673-789-9  
; Sequence 9, Application US/08673789  
; Patent No. 5814479  
; GENERAL INFORMATION:  
; APPLICANT: ZHOU, RENPING; SCHULZ, NICHOLAS,  
; APPLICANT: T.; KROMER, LAWRENCE, F.; VANDE WOUDE,  
; APPLICANT: GEORGE, F.  
; TITLE OF INVENTION: BSK RECEPTOR LIKE  
; TITLE OF INVENTION: TYROSINE KINASE AND LIGAND AND THEIR  
; TITLE OF INVENTION: USE IN DIAGNOSTIC AND THERAPEUTIC  
; TITLE OF INVENTION: METHODS  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORGAN & FINNEGAN  
; STREET: 345 PARK AVENUE  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: USA  
; ZIP: 10154  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: FLOPPY DISK  
; COMPUTER: IBM PC COMPATIBLE  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WORDPERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/673,789  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/177,812  
; FILING DATE: 04-JAN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CAROL M. GRUPPI  
; REGISTRATION NUMBER: 37,341  
; REFERENCE/DOCKET NUMBER: 2026-4105  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 758-4800  
; TELEFAX: (212) 751-6849  
; TELETYPE: 421792  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 984  
; TYPE: AMINO ACID  
; STRANDEDNESS: UNKNOWN  
; TOPOLOGY: UNKNOWN  
; US-08-673-789-9  
Query Match 10.3%; Score 73.5; DB 2; Length 984;  
Best Local Similarity 27.5%; Pred. No. 2;

	Matches	36:	Conservative	20:	Mismatches	52:	Indels	23:	Gaps	6:
Qy	7	RVERFLQVFSAGCKT-----PIHLKDCVCALYNFNEDEEAVALFEPQHSDLLHCRITIE	60.							
		:	:	:	:	:	:	:	:	:
Db	94	RHVHELTFTVRDCKSFPGGAGPLCKETFNLULYMESQDVGI----QLRRFLPKQVTVA	149							
		:    :	:	:	:	:	:	:	:	:
Qy	61	ADPOTSI-TLYSMLLQLNFENAAAM-----RCWCWALDE-----LHNVRCLCFQOSLEHD	108							
		:    :    :	:	:	:	:	:	:	:	:
Db	150	AQQSFTTRDLASGSVKLNVEHCISGLRITRGRLYLAFHPNPGACVALSVRVYQRCPETLN	209							
		:    :    :	:	:	:	:	:	:	:	:
Qy	109	E-ASFSDIVSG	118							
		:	:	:	:	:	:	:	:	:
Db	210	GLAQFPDTLP	220							
		:	:	:	:	:	:	:	:	:

```

RESULT      4
US-08-449-645A-19
; Sequence 19, Application US/08449645A
; Patent No. 5981245
; GENERAL INFORMATION:
; APPLICANT: Fox, Gary M.
; TITLE OF INVENTION: EPH-Like Receptor Protein Tyrosine
; TITLE OF INVENTION: Kinases
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Patent Operations/RBW
; STREET: 1840 Behavilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/449,645A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Winter, Robert B.
; REFERENCE/DOCKET NUMBER: A-287
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 984 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-449-645A-19

```

Query Match	10.3%;	Score 73.5;	DB 2;	Length 984;	
Best Local Similarity	27.5%;	Pred. No. 2;			
Matches	36;	Conservative 20;	Mismatches 52;	Indels 23;	Gaps 6;
2Y	7	RVERLQYFESACKT-----PHLLKDGVCALYNEDDEAAVL	EVPOHSDSLLLHCRRIE	60	
		:	:	:	:
db	94	RVHVELQFTVRDCKFPGGAGPLGKRETNLLYMESDQDVGI---	QLRRPLFKQVTTVA	149	
		:	:	:	:
2Y	61	ADPQTSI-TLYSMLQLQNFEMAAM-----RGCMWALDE-----	LHNVRLCFQGSLEHLD	108	
		:	:	:	:
db	150	ADQSFTRDLASGVKLNVERCSLGRLTRGLYAFHNFACVALSVRVFYRCPELTN		209	
		:	:	:	:
2Y	109	E-ASFSDIVSG	118		
		:	:	:	:
db	210	GLAQFPDPTLPG	220		
		:	:	:	:

RESULT 5  
US-08-702-367A-19  
Sequence 19, Application US/08702367A

```

; Patent No. 5981246
; GENERAL INFORMATION:
; APPLICANT: Fox, Gary M.
; TITLE OF INVENTION: EPH-Like Receptor Protein Tyrosine
; TITLE OF INVENTION: Kinases
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Patent Operations/RBM
; STREET: 1840 Behavilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/702,367A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Winter, Robert B.
; REFERENCE/DOCKET NUMBER: A-287
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 984 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-702-367A-19

```

	Query Match	10.3%;	Score 73.5;	DB 2;	Length 984;	
	Rest Local Similarity	27.5%;	Pred. No. 2;			
	Matches	36;	Conservative	20;	Mismatches	52; Indels 23; Gaps
QY	7 RVERFLOYFSAGCKT-----PIHLKDCVCALYNQDEEEAAVLEVPQHSDSLLLHCRIIE	60				
	:	:	:	:	:	:
Dd	94 RVHVELOFTVRDCKSPFGGACPLGCKETFNLLLYMESQDVGI----QLRRPFLFKVTVA	149				
	:     :	:	:	:	:	:
QY	61 ADPQTST-TLYSMLLQNLFEMAAM-----RCGWLALDE-----LHNVRCLCFOQSLEHLD	108				
	:    :     :	:	:	:	:	:
Dd	150 AQDSFTIRDLASGSVKLNVERCSIGRLTRGLYLAFHPNPGACVALSVRVRYFQRCPETLN	209				
	:     :	:	:	:	:	:
QY	109 E-ASFSDIVSG	118				
	:	:	:	:	:	:
Dd	210 GLAQFFDTLP	220				

```

RESULT      6
PCT-US95-04681-19
: Sequence 19, Application PC/TUS9504681
: GENERAL INFORMATION:
: APPLICANT: Fox, Gary M.
: TITLE OF INVENTION: Eph-Like Receptor Protein Tyrosine
: : Kinases
: NUMBER OF SEQUENCES: 28
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Amgen Patent Operations/RBW
: STREET: 1840 Dehavilland Drive
: CITY: Thousand Oaks
: STATE: California
: COUNTRY: USA
: ZIP: 91320
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:

```

TELEFAX: 703-685-0573  
TELEX: 248425 EMBON  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2183 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-348-891A-7

Query Match 9.7%; Score 69.5; DB 1; Length 2183;  
Best Local Similarity 20.0%; Pred. No. 20;  
Matches 38; Conservative 26; Mismatches 63; Indels 63; Gaps 7;

QY 4 SQQRVERFLQYFSAGCK-----TPHLKDGVCALYNEQDEAAVLEVP 46  
DB 158 SSQWFEPFLWFTVKTMRSVTKSQTHTCHRRRHPVFFTGSSVELLSRLDLVAISKES 217  
QY 47 QH-----SDSLLLHCRRIIEAD--PQTSITLYSMLQLNFEMAMRCGWLALD----- 91  
DB 218 QHVVYLTFLVLMYCDVIEGRLMTETAMTIDARYTEL---LGRVRYMWKLLIDGFFPALGN 274  
QY 92 -----ELHNVRL--CFOOSLEHLDEASFSDIVSGFIEHAAE 125  
DB 275 PTYQIVAMLEPLSLAYLQLRDITVELRGAFNLHCFTFIEHVDLDQNGFSD--EGTYHELIE 332  
QY 126 VREYIAQLDE 135  
DB 333 ALDYIFITDD 342

RESULT 8  
US-08-905-817-7  
Sequence 7, Application US/08905817  
Patent No. 5824777  
GENERAL INFORMATION:  
APPLICANT: SASAKI, Keiko  
APPLICANT: MORI, Takayuki  
TITLE OF INVENTION: ATTENUATED MEASLES VIRUS VACCINE,  
CONTAINING SPECIFIC NUCLEOTIDE SEQUENCE AND A METHOD FOR  
ITS ABSOLUTE IDENTIFICATION  
TITLE OF INVENTION: ITS ABSOLUTE IDENTIFICATION  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: YOUNG & THOMPSON  
STREET: 745 South 23rd Street  
CITY: Arlington  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/905,817  
FILING DATE: 04-AUG-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/348,891  
FILING DATE: 25-NOV-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/848,400  
FILING DATE: 10-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 3-293625  
FILING DATE: 14-OCT-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: PATCH, Andrew J.  
REGISTRATION NUMBER: 32,925  
REFERENCE/DOCKET NUMBER: KP-7501A

APPLICATION NUMBER: PCT/US95/04681  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Winter, Robert B.  
REFERENCE/DOCKET NUMBER: A-287  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 984 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US95-04681-19

Query Match 10.3%; Score 73.5; DB 5; Length 984;  
Best Local Similarity 27.5%; Pred. No. 2;  
Matches 36; Conservative 20; Mismatches 52; Indels 23; Gaps 6;

QY 7 RVERFLQYFSAGCKT-----PIHLKDGVCALYNEQDEAAVLEVPQHSLSLLHCRILIE 60  
DB 94 RVHVELOFTVRDCKSPGGAGPLGCKETFNLLYMESDQDVGI-----QLRRPLFKQVTTVA 149  
QY 61 ADPQTSI-TLYSMLQLNFEMAM-----RCQWLALDE-----LHNVRLCFQOSLEHL 108  
DB 150 ADQSFTRIDIASGVKLNVERCSGLRTRRGLYLAFHNPACVALVSVRYFYQRCPETLN 209  
QY 109 E-ASFSDIVSG 118  
DB 210 GLAQFPDTLPS 220

RESULT 7  
US-08-348-891A-7  
Sequence 7, Application US/08348891A  
Patent No. 5654136  
GENERAL INFORMATION:  
APPLICANT: SASAKI, Keiko  
APPLICANT: MORI, Takayuki  
TITLE OF INVENTION: ATTENUATED MEASLES VIRUS VACCINE,  
CONTAINING SPECIFIC NUCLEOTIDE SEQUENCE AND A METHOD FOR  
ITS ABSOLUTE IDENTIFICATION  
TITLE OF INVENTION: ITS ABSOLUTE IDENTIFICATION  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: YOUNG & THOMPSON  
STREET: 745 South 23rd Street  
CITY: Arlington  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/348,891A  
FILING DATE: 25-NOV-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/848,400  
FILING DATE: 10-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 3-293625  
FILING DATE: 14-OCT-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: PATCH, Andrew J.  
REGISTRATION NUMBER: 32,925  
REFERENCE/DOCKET NUMBER: KP-7501  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-521-2297



[illegible]

```

RESULT 11
US-08-628-145-5
; Sequence 5, Application US/08628145
; Patent No. 5872214
; GENERAL INFORMATION:
; APPLICANT: Seizinger, Bernd R.
; APPLICANT: Kleyer, Nikolai A.
; APPLICANT: Bianchi, Albert B.
; TITLE OF INVENTION: No. 5872214el NF2 Isoforms
; NUMBER OF SEQUENCES: 26

```

CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Read & Robins  
 STREET: 635 Bryant Street  
 CITY: Palo Alto  
 STATE: California  
 COUNTRY: U.S.A  
 ZIP: 94301  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/628,145  
 FILING DATE: 04-APR-1996

CLASSIFICATION: 530  
PRIOR APPLICATION DATA: US 08/179,738  
FILING DATE: 10-JAN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Robins, Roberta L.  
REGISTRATION NUMBER: 33,208  
REFERENCE/DOCKET NUMBER: 5998-0017  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 617-8999  
TELEFAX: (415) 327-3231  
INFORMATION FOR SEQ ID NO: 5:

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 591 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: murine
; US-08-628-145-5

```

Query Match	9.6%	Score 68.5;	DB 2;	Length 591;
Best Local Similarity	25.7%	Pred. No. 4;		
Matches	27; Conservative	22; Mismatches	43; Indels	1

```

QY      33 YNEQDEEAAVLEVPQPHSDSLLHCRITE----ADPQTSITLYSMLQL---NFEAAMRG 85
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     101 YPENAEELVQEIQTHLFFLVQKKQILDEKVKCPPEASVILLASTAVAKYGDYDPFSVHKR 160

```

QY 86 CWLALDELHNVL--CFQSQLEHLDEASFSDIVSGFIEHAAEVPRE 128  
          ::||::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:  
Db 161 GFLAQEEELLPKRVINYQTPEWMEER-----ITAWVAEHGRGARD 201

RESULT 12  
US-08-628-145-10  
; Sequence 10, Application US/08628145  
; Patent No. 5972114  
; GENERAL INFORMATION:  
; APPLICANT: Seizinger, Bernd R.  
; APPLICANT: Kley, Nikolai A.  
; APPLICANT: Bianchi, Albert B.

```

1  TITLE OF INVENTION:  No. 5872214el NF2 Isoforms
2
3  NUMBER OF SEQUENCES: 26
4  CORRESPONDENCE ADDRESS:
5  ADDRESSEE:  Reed & Robins
6  STREET:  635 Bryant Street
7  CITY:  Palo Alto
8  STATE:  California
9  COUNTRY:  U.S.A
10
11  ZIP:  94301
12
13  COMPUTER READABLE FORM:
14  MEDIUM TYPE:  Floppy disk
15  COMPUTER:  IBM PC compatible
16  OPERATING SYSTEM:  PC-DOS/MS-DOS
17  SOFTWARE:  Patentin Release #1.0, Version #1.25
18
19  CURRENT APPLICATION DATA:
20  APPLICATION NUMBER:  US/08/628,145
21  FILING DATE:  04-APR-1996
22  CLASSIFICATION:  530
23
24  PRIOR APPLICATION DATA:
25  APPLICATION NUMBER:  US 08/179,738
26  FILING DATE:  10-JAN-1994
27  ATTORNEY/AGENT INFORMATION:
28  NAME:  Robins, Roberta L.
29  REGISTRATION NUMBER:  33,208
30  REFERENCE/DOCKET NUMBER:  5998-0017
31
32  TELECOMMUNICATION INFORMATION:
33  TELEPHONE:  (415) 617-8999
34  TELEFAX:  (415) 327-3231
35
36  INFORMATION FOR SEQ ID NO: 10:
37  SEQUENCE CHARACTERISTICS:
38  LENGTH: 591 amino acids
39  TYPE: amino acid
40  TOPOLOGY: linear
41
42  MOLECULE TYPE: protein
43  FRAGMENT TYPE: N-terminal
44  ORIGINAL SOURCE:
45  ORGANISM: Homo sapien
46
47  US-08-628-145-10

```

Query Match	9.6%	Score 68.5	DB 2	Length 591	
Best Local Similarity	25.7%	Prod. No. 4			
Matches	27	Conservative	22	Mismatches 43	
				Indels 13	
Qy	33	YNEQDEEAAVLEVPQSDLSLLHCRRIIE	---	ADPQTSTITLYSMLLQL---	NFEMAF
Db	101	YPENAEELVQELTHLFFLVQVKQIILDEKTYCPPEASVLASVQAQYKGYDPS			
Qy	86	CWLALDELHNVRL	-	CFQQSLHLEHDEASFSDIVSGFIEHAAEVRE	128
Db	161	GFLAQEELLPKRVINLYQMTPEMWEER	----	ITAWYAEHGRGRAD	201

RESULT 13  
US-08-171-718-16  
; Sequence 16, Application US/08171718  
; Patent No. 5707863

```

: GENERAL INFORMATION:
:
: APPLICANT: Trofatter, James A.
: APPLICANT: MacCollin, Mia M.
: APPLICANT: Gusella, James F.
: TITLE OF INVENTION: Tumor Suppressor Gene Merlin and Uses
: TITLE OF INVENTION: Thereof
: NUMBER OF SEQUENCES: 120
: CORRESPONDENCE ADDRESS:
: ADDRESSER: Sterne, Kessler, Goldstein & Fox
: STREET: 1100 New York Avenue, N.W., Suite 600
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20005-3934
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk

```



COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/171,718  
FILING DATE: 22-DEC-1993  
CLASSIFICATION: 436  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/108,808  
FILING DATE: 19-AUG-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/022,034  
FILING DATE: 25-FEB-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/026,063  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Brown, Anne  
REGISTRATION NUMBER: 36,463  
REFERENCE/DOCKET NUMBER: 0609.3850003  
TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 595 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-171-718-16

Query Match 9.6%; Score 68.5; DB 1; Length 595;  
Best Local Similarity 25.7%; Pred. No. 4.1;  
Matches 27; Conservative 22; Mismatches 43; Indels 13; Gaps 4;  
QY 33 YNEQDEAAVLEVPQHSLSLLHCRIE----ADPQTSITLYSMLLQL---NFEMAAAMRG 85  
DB 101 YPENAEELVQEIQTQHLFFLOVKKQILDEKIYCPPEASVLLASVAVQAKYGDYDPSVHKR 160  
QY 86 CWLALDELHNRL--CFQOSLEHLDEASFSDIIVSGFIEHAAEVRE 128  
DB 161 GFLAQEELLPKRVINLYQMTPEMWEER---ITAWYAEHRGRARD 201

RESULT 14  
US-08-478-087-16  
Sequence 16, Application US/08478087  
Patent No. 6077685  
GENERAL INFORMATION:  
APPLICANT: Trofatter, James A.  
APPLICANT: MacCollin, Mia M.  
APPLICANT: Gusella, James F.  
TITLE OF INVENTION: Tumor Suppressor Gene Merlin and Uses  
TITLE OF INVENTION: Thereof  
NUMBER OF SEQUENCES: 120  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
STREET: 1100 New York Avenue, N.W., Suite 600  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3934  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/478,087  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/171,718  
FILING DATE: 22-DEC-1993  
APPLICATION NUMBER: US 08/108,808  
FILING DATE: 19-AUG-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/022,034  
FILING DATE: 25-FEB-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/026,063  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Brown, Anne  
REGISTRATION NUMBER: 36,463  
REFERENCE/DOCKET NUMBER: 0609.3850003  
TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 595 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-478-087-16

Query Match 9.6%; Score 68.5; DB 3; Length 595;  
Best Local Similarity 25.7%; Pred. No. 4.1;  
Matches 27; Conservative 22; Mismatches 43; Indels 13; Gaps 4;  
QY 33 YNEQDEAAVLEVPQHSLSLLHCRIE----ADPQTSITLYSMLLQL---NFEMAAAMRG 85  
DB 101 YPENAEELVQEIQTQHLFFLOVKKQILDEKIYCPPEASVLLASVAVQAKYGDYDPSVHKR 160  
QY 86 CWLALDELHNRL--CFQOSLEHLDEASFSDIIVSGFIEHAAEVRE 128  
DB 161 GFLAQEELLPKRVINLYQMTPEMWEER---ITAWYAEHRGRARD 201

RESULT 15  
US-08-179-738-2  
Sequence 2, Application US/08179738  
Patent No. 5578462  
GENERAL INFORMATION:  
APPLICANT: Seizinger, Bernd R.  
APPLICANT: Kleiz, Nikolai A.  
APPLICANT: Bianchi, Albert B.  
TITLE OF INVENTION: NO. 5578462el NF2 Isoforms  
NUMBER OF SEQUENCES: 26  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Reed & Robins  
STREET: 635 Bryant Street  
CITY: Palo Alto  
STATE: California  
COUNTRY: U.S.A  
ZIP: 94301  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/179,738  
FILING DATE: 10-JAN-1994  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Robins, Roberta L.  
REGISTRATION NUMBER: 33,208  
REFERENCE/DOCKET NUMBER: 5998-0017  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 617-8999  
TELEFAX: (415) 327-3231  
INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:  
; LENGTH: 596 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FRAGMENT TYPE: N-terminal  
; ORIGINAL SOURCE:  
; ORGANISM: murine  
; US-08-179-738-2

Query Match 9.6%; Score 68.5; DB 1; Length 596;  
Best Local Similarity 25.7%; Pred. No. 4.1;  
Matches 27; Conservative 22; Mismatches 43; Indels 13; Gaps 4;  
Qy 33 YNEQDEEAALVLPQHSLSLLHCRIE---ADPQTSITILYSMLLQL---NFEMAAAMRG 85  
Db 101 YPENAEELVQETITQHLFFLQVKKKQILDEKVCPEASVLLASAYQAKYGDYDPSVHKR 160  
Qy 86 CWLALDELHNRL--CFQGSLEHLDEASFSDIVSGFIEHAAEVRE 128  
Db 161 GFLLAQEELLPKRVINLYQMTPEMWEER---ITAWVAEHRGRARD 201

Search completed: June 5, 2001, 18:17:00  
Job time: 186 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2001, 18:16:36 ; Search time 55.68 Seconds  
(without alignments)  
142.703 Million cell updates/sec

Title: US-09-596-784-4  
Perfect score: 715  
Sequence: 1 MTSSQQRVERFLQFSAGCK.....IEHAAEVREYIAQLDESSAA 139

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_0401.\*  
1: /SID56/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SID56/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SID56/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SID56/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SID56/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SID56/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SID56/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SID56/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SID56/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SID56/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SID56/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SID56/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SID56/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SID56/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SID56/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SID56/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SID56/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SID56/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SID56/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SID56/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SID56/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SID56/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	715	100.0	139	W98012	Hypersensitive res
2	715	100.0	139	PA	Erwinia amylovora
3	715	100.0	139	Y71096	A hypersensitive r
4	80.5	11.3	368	W89180	Amino acid sequenc
5	80.5	11.3	766	B07739	A snake venom prot
6	80.5	11.3	787	B07740	A snake venom prot
7	80.5	11.3	820	B07741	A snake venom prot
8	76.5	10.7	1105	W44864	Human TPC2 telomer
9	76.5	10.7	1105	W73958	Human TPC2 protein
10	75	10.5	243	W38542	Streptococcus pneu
11	75	10.5	317	Y85961	S. pneumoniae deri

12	75	10.5	317	21	B13516
13	72.5	10.1	466	22	B63634
14	71.5	10.0	920	19	W82500
15	70.5	9.9	2183	19	W48708
16	70.5	9.9	2183	19	W48704
17	69.5	9.7	2183	14	R39592
18	69.5	9.7	2183	19	W48706
19	69.5	9.7	2183	19	W48707
20	69.5	9.7	2183	19	W48709
21	69.5	9.7	2183	19	W48710
22	69.5	9.7	2183	19	W48703
23	69.5	9.7	2183	19	W48705
24	68.5	9.6	239	20	Y25380
25	68.5	9.6	239	20	Y02636
26	68.5	9.6	584	20	W94457
27	68.5	9.6	590	20	W94458
28	68.5	9.6	591	18	W09646
29	68.5	9.6	591	18	W09648
30	68.5	9.6	591	20	W94456
31	68.5	9.6	595	15	R60398
32	68.5	9.6	596	18	W09645
33	68.5	9.6	596	20	W94455
34	68.5	9.6	596	20	W94459
35	68	9.5	225	20	Y34662
36	68	9.5	652	17	R88124
37	68	9.5	1087	20	Y19935
38	68	9.5	1119	20	Y19934
39	68	9.5	1144	17	R88122
40	68	9.5	1144	17	R88123
41	67.5	9.4	404	21	B20940
42	67.5	9.4	584	18	W09647
43	67.5	9.4	1527	20	Y43543
44	67.5	9.4	1528	18	W33363
45	67	9.4	1194	21	B36533

#### ALIGNMENTS

RESULT 1  
W98012  
ID W98012 standard; Protein; 139 AA.  
XX  
AC W98012;  
XX  
DT 21-JUN-1999 (first entry)  
XX  
DE Hypersensitive response elicitor DspF.  
DE  
DE Hypersensitive response elicitor; DspF; disease resistance;  
KW Insect resistance; biological control; transgenic plant.  
XX  
OS Erwinia amylovora.  
XX  
PN W09907206-A1.  
XX  
PD 18-FEB-1999.  
XX  
PF 24-JUL-1998; 98WO-US15426.  
XX  
PR 06-AUG-1997; 97US-0055105.  
XX  
XX (CORR ) CORNELL RES FOUND INC.  
XX  
XX Beer SV, Bogdanove AJ, Kim JF, Wei Z;  
XX WPI; 1999-180362/15.  
XX N-PSDB; X24811.  
XX  
XX Nucleic acid encoding hypersensitive response-eliciting protein -  
XX used to improve growth of plants and impart resistance to disease  
XX and insects

Streptococcus pneu  
Human gastric canc  
Human OCT protein.  
Measles virus Mora  
Measles virus 1977  
L protein of atten  
Measles virus Mont  
Measles virus Rube  
Measles virus Zagr  
Measles virus Aik-  
Measles virus Edmo  
HPV fusion protein  
CLYTA-E7-His prote  
Mouse neurofibroma  
Human neurofibroma  
Mouse merlin prote  
Human merlin prote  
Mouse neurofibroma  
Merlin protein enc  
Mouse merlin prote  
Mouse neurofibroma  
Human neurofibroma  
Chlamydia pneumoni  
Tobacco mosaic vir  
B. burgdorferi ant  
B. burgdorferi ant  
Tobacco mosaic vir  
Drosophila odorant  
Mouse merlin prote  
A human MPR-relate  
Human multidrug re  
Preliminary CLASP-

XX PI Wei Z, Schading RL;  
 XX DR WPI; 2000-376566/32.  
 XX DR N-PSDB; D00671.  
 XX PT Application of a hypersensitive response elicitor protein to plants to  
 XX PT impart stress resistance  
 XX PS Disclosure; Page 21; 84pp; English.  
 XX CC The patent discloses a method to impart stress resistance to plants by  
 XX CC applying a hypersensitive response elicitor in a non-infectious form to  
 XX CC a plant or seed. The present sequence is a hypersensitive response  
 XX CC elicitor encoded by dspF gene from *Erwinia amylovora*.  
 XX CC The protein is used to impart stress resistance to plants.  
 XX SQ Sequence 139 AA;  
 Query Match 100.0%; Score 715; DB 21; Length 139;  
 Best Local Similarity 100.0%; Pred. No. 9e-79;  
 Matches 139; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MTSSQORVERFLOYFSAGCKTPHILKDGVCALYNEQDEAAVLEVPQHSLSLLHCRRIE 60  
 DB 1 mtssqqrverflqyfsagcktpihikdgvcalyneqdeaaavlevpqhsllhcrrie 60  
 QY 61 ADPQTSITLYSMLQLNFEMAMRGCVLALDELHNVRCLCFQOSLEHLDEASFSDIVSGFI 120  
 DB 61 adpqtstilysmllqlnfemaamrgcwlaldelhnvrclcfqoslehldeasfsdivsgfi 120  
 QY 121 EHAAEVREYIAQLDESSAA 139  
 DB 121 ehaaevreyiaqldessaa 139  
 RESULT 3  
 Y84857  
 ID Y84857 standard; Protein; 139 AA.  
 XX AC Y84857;  
 XX DT 08-AUG-2000 (first entry)  
 XX DE A hypersensitive response elicitor protein.  
 XX KW Hypersensitive response; insect control; disease resistance;  
 XX KW hypersensitive response elicitor; plant growth; vegetable; crop;  
 XX KW ornamental plant; dspF gene.  
 XX OS *Erwinia amylovora*.  
 XX PN WO200020452-A2.  
 XX PD 13-APR-2000.  
 XX PF 05-OCT-1999; 99WO-US23181.  
 XX PR 05-OCT-1998; 98US-0103050.  
 XX PA (EDEN-) EDEN BIOSCIENCE CORP.  
 XX PI Wei Z, Fan H, Niggemeyer JL;  
 XX DR WPI; 2000-303745/26.  
 XX DR N-PSDB; A14941.  
 XX PT Hypersensitive response elicitor polypeptides useful for imparting  
 XX PT enhanced growth, disease resistance and insect resistance to plants,  
 XX PT especially vegetables and ornamental flowers -  
 XX PS Disclosure; Page 22-23; 100pp; English.

PS Claim 18; Page 57-58; 75pp; English.  
 XX CC This polypeptide comprises the 16 kDa hypersensitive response  
 XX CC elicitor protein DspF of *Erwinia amylovora*. The nucleotide  
 XX CC sequence of the dsp region of *E. amylovora* strain Ea321 was  
 XX CC determined using subclones of pCPP430. A two-gene operon was  
 XX CC discovered comprising dspe (see X24810) and dspF (see X24811).  
 XX CC isolated dsp DNA molecules and encoded proteins can be used to  
 XX CC impart disease resistance to plants, to enhance plant growth,  
 XX CC and/or to control insects on plants. This is achieved by applying  
 XX CC a hypersensitive response elicitor protein or polypeptide in a  
 XX CC non-infectious form to plants or plant seeds, or by producing a  
 XX CC transgenic plants or plant seeds transformed with DNA encoding a  
 XX CC hypersensitive response elicitor. Protection can be provided  
 XX CC against a wide range of viruses, bacteria, fungi and insects, e.g.  
 XX CC tobacco mosaic virus and tomato mosaic virus, *Pseudomonas syringae*,  
 XX CC *Xanthomonas campestris*, *Fusarium oxysporum*, *Phytophthora infestans*,  
 XX CC armyworm, diamondback moth, etc. The method avoids use of  
 XX CC infectious agents or polluting chemicals. Claimed transgenic  
 XX CC plants are selected from alfalfa, rice, wheat, barley, rye, cotton,  
 XX CC sunflower, peanut, corn, potato, bean, pea, chichory, lettuce,  
 XX CC endive, cabbage, brussel sprout, sweet potato, beet, parsnip,  
 XX CC turnip, cauliflower, broccoli, turnip, radish, spinach, onion,  
 XX CC garlic, eggplant, pepper, celery, carrot, squash, pumpkin,  
 XX CC zucchini, cucumber, apple, pear, melon, citrus, strawberry, grape,  
 XX CC raspberry, pineapple, soybean, tobacco, tomato, sorghum, sugarcane,  
 XX CC *Arabidopsis thaliana*, *Sainpaulia*, *Petunia*, *Pelargonium*, *Poinsettia*,  
 XX CC *Chrysanthemum*, *Carnation* and *Zinnia*.  
 XX SQ Sequence 139 AA;  
 Query Match 100.0%; Score 715; DB 20; Length 139;  
 Best Local Similarity 100.0%; Pred. No. 9e-79;  
 Matches 139; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MTSSQORVERFLOYFSAGCKTPHILKDGVCALYNEQDEAAVLEVPQHSLSLLHCRRIE 60  
 DB 1 mtssqqrverflqyfsagcktpihikdgvcalyneqdeaaavlevpqhsllhcrrie 60  
 QY 61 ADPQTSITLYSMLQLNFEMAMRGCVLALDELHNVRCLCFQOSLEHLDEASFSDIVSGFI 120  
 DB 61 adpqtstilysmllqlnfemaamrgcwlaldelhnvrclcfqoslehldeasfsdivsgfi 120  
 QY 121 EHAAEVREYIAQLDESSAA 139  
 DB 121 ehaaevreyiaqldessaa 139  
 RESULT 2  
 Y71096  
 ID Y71096 standard; Protein; 139 AA.  
 XX AC Y71096;  
 XX DT 08-SEP-2000 (first entry)  
 XX DE *Erwinia amylovora* hypersensitive response elicitor encoded by dspF gene.  
 XX KW Hypersensitive response elicitor; environmental stress resistance;  
 XX KW plant; dspF gene.  
 XX OS *Erwinia amylovora*.  
 XX PN WO200028055-A2.  
 XX PD 18-MAY-2000.  
 XX PF 04-NOV-1999; 99WO-US26039.  
 XX PR 05-NOV-1998; 98US-0107243.  
 XX PA (EDEN-) EDEN BIOSCIENCE CORP.

XX The present sequence represents a hypersensitive response elicitor  
 CC polypeptide. The polynucleotide represents the dsf gene. The  
 CC specification describes hypersensitive response elicitor polypeptide  
 CC fragments, which do not elicit a hypersensitive response. Instead,  
 CC the proteins impart disease resistance to plants. Instead,  
 CC growth, and/or control insects. The polypeptide fragments may be  
 CC used to these properties to plants. The plants which may be treated  
 CC in this way include vegetables, crops and ornamental plants such as  
 CC alfalfa, rice, wheat, barley, rye, cotton, sunflower, peanut, corn,  
 CC potato, sweet potato, bean, pea, chickory, lettuce, endive, cabbage,  
 CC brussel sprout, beet, parsnip, turnip, cauliflower, broccoli, radish,  
 CC spinach, onion, garlic, eggplant, pepper, celery, carrot, squash,  
 CC pumpkin, zucchini, cucumber, apple, pear, melon, citrus, strawberry,  
 CC grape, raspberry, pineapple, soybean, tobacco, tomato, sorghum or  
 CC sugarcane, Arabidopsis thaliana, Saintpaulia, petunia, pelargonium,  
 CC poinsettia, chrysanthemum, carnation or zinnia.  
 XX Sequence 139 AA;

Query Match 100.0%; Score 715; DB 21; Length 139;  
 Best Local Similarity 100.0%; Pred. No. 9e-79;  
 Matches 139; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MTSSQQRVERFLQYFSAGCKTPHLKDGVCALYNEQDEEAALVLPQHSLSLHCRITIE 60  
 DB 1 mtssqqrverflqyfsagcktphlkdgvcallyneqdeeaavlevpqhsdlsllhcrille 60  
 QY 61 ADPQTSITLYSMLOLNFEAMRGCVLWALDELHNRLCFQOQSLHLEDEASFSDIVSGFI 120  
 DB 61 adpqtstilysmllqnfemamrgcwlaldelhnrclfcqgslhlehdeasfsdivsgfi 120  
 QY 121 EHAEEVREYIAQLDESSAA 139  
 DB 121 ehaeavreyiaqldeessaa 139

RESULT 4  
 ID W89180 standard; Protein; 368 AA.  
 AC W89180;  
 DT 11-MAR-1999 (first entry)  
 DE Amino acid sequence of delta9 14:0-ACP desaturase.  
 KW delta 9 14:0-ACP desaturase; fatty acid; anacardic acid; pest;  
 KW resistance; plant; pelargonium; geranium; polymer; nylon.  
 OS Pelargonium sp.  
 PN US5856157-A.  
 PD 05-JAN-1999.  
 XX 04-JUN-1997; 97US-0869137.  
 XX 04-JUN-1996; 96US-0018957.  
 PR 04-JUN-1997; 97US-0869137.  
 XX (PENN-) PENN STATE RES FOUND.  
 PA Cox-foster DL, Craig R, Medford JI, Mumma RO, Schultz D;  
 PI WPI: 1999-105114/09.  
 DR N-PSDB; V81284.  
 XX New isolated 14:0-ACP desaturase gene - obtained from geranium  
 PT plants, used to enhance pest resistance of plants and for enhancing  
 PT production of unsaturated fatty acids in plants  
 XX

PS Claim 1; Fig 3; 26pp; English.

XX This represents a delta9 14:0-ACP desaturase. The 14:0-ACP desaturase  
 CC gene enhances the accumulation 16:1 Delta-11 and 18:1 Delta 3 fatty acids  
 CC and 22:1 omega 5 and 24:1 omega 5 anacardic acids. The gene can be used  
 CC to provide pest resistance in plants. The gene can also be used to  
 CC enhance the production of unsaturated fatty acids in plants such as  
 CC soybeans, rapeseed, maize, sunflower, safflower, cotton, cuphea, peanut,  
 CC coconut, oil-palm and pelargonium. The gene and its expression products  
 CC can also be used in manufacturing methods, e.g. in innovative processes  
 CC for making specialty polymers such as nylon and other methods in which  
 CC unsaturated fatty acids are used as constituents or starting materials.  
 CC In addition the promoter from the gene can be used for trichome specific  
 CC expression of polypeptides.

SQ Sequence 368 AA;

Query Match 11.3%; Score 80.5; DB 20; Length 368;  
 Best Local Similarity 20.1%; Pred. No. 0.24;  
 Matches 37; Conservative 26; Mismatches 54; Indels 67; Gaps 8;  
 QY 6 QRVERFLQYFSA-----GC-KTPIHLKDGVCALYNEQDEEAALVLE-----VPOHSDSL 52  
 DB 169 rqvextiqylialgqdgiteknpvhl-----fiytsfqratfishantaklaqghdkq 223  
 QY 53 LLH-CRITTEADPQTSITLYSMLOLNFE-----fiytsfqratfishantaklaqghdkq 80  
 DB 224 laiqcgtlaadekrhetayirivdklfeldpdtmsclahmkrkrtmpahlmrdgrdph 283  
 QY 81 -----AAMRGCVLWALDELHNRLCFQOQSLH-LDEASFSDIVSGFIEHAEEVREYIA 131  
 DB 284 lfqhfsvvartgvtyvmdyin-----ilehfvkwniekitaglsdkgreagdyvc 335  
 QY 132 QLDE 135  
 DB 336 klge 339

RESULT 5  
 B07739  
 ID B07739 standard; Protein; 766 AA.  
 AC B07739;  
 DT 07-NOV-2000 (first entry)  
 DE A snake venom protease (SVPH-1) polypeptide variant SVPH-1a.  
 KW Snake venom protease; SVPH; SVPH-1; metalloproteinase-disintegrin;  
 KW chromosome 1; chromosome 4; immune system; splice variant.  
 OS Homo sapiens.  
 PN WO200043525-A2.  
 PD 27-JUL-2000.  
 XX 21-JAN-2000; 2000WO-US01338.  
 PF 21-JAN-1999; 99US-0116670.  
 PR 14-JUN-1999; 99US-0138682.  
 PR 27-SEP-1999; 99US-0155798.  
 XX (IMMV) IMMUNEX CORP.  
 PA Cerretti DP;  
 PI WPI: 2000-482914/42.  
 DR N-PSDB; A59304.  
 XX Snake venom protease (SVPH) nucleic acids, and polypeptides, used to  
 PT identify proteins having metalloproteinase-disintegrin activity, and

inhibitors of the proteins for use in therapeutics -

Claim 13; Page 15; 105pp; English.

The present sequence represents an alternatively spliced snake venom protease-1 (SVPH-1) polypeptide, designated SVPH-1a. The SVPH polypeptides are metalloproteinase-disintegrin protein family members. The SVPH polynucleotides can be used as probes to identify nucleic acids encoding proteins having metalloproteinase-disintegrin activity, to identify human chromosome 1 or 4, to map genes on those chromosomes, to identify genes associated with diseases, syndromes and conditions associated with the chromosomes, and to study proteinases and their activities on cell/cell interactions and the immune system. Sense or antisense oligonucleotides of SVPH can be used to inhibit gene expression of SVPH 1, 3, or 4. The SVPH polypeptides can be used to study cell/cell and cell/matrix interactions involved in cellular processes and in the immune system. The polypeptides may also be used to screen for inhibitors of the polypeptide's activity, which are used in therapeutics. The antibodies can be used in assays to detect the presence of the polypeptides in vitro or in vivo, and to purify the polypeptides by affinity chromatography.

XX Sequence 766 AA;

Query Match 11.3%; Score 80.5; DB 21; Length 766;

Best Local Similarity 22.3%; Pred. No. 0.67; 47; Indels 33; Gaps 6;

Matches 31; Conservative 28; Mismatches 28;

QY 18 GCKTPHILK-----DGVCALYNQDEFAAVLEVPQHSLSLLHCRIEADPQTSITL-- 69

Db 60 gqkhihiakvkkllfshlpvftydqgalledqpfvqncyyh-gyvegdpeslvslst 118

QY 70 ----YSMLLQLNFEMAAWGCLALDELHNVR-LCFOQSLEHL-----DEASFSDIVSG 118

Db 119 cfggfgilqin-----dfayeikplafstfthfvlvykmdseekqfstmrg 165

QY 119 FIEHAAVEHYIAQLDESS 137

Db 166 fmqneitcrmeifeidnst 184

RESULT 6

B07740

ID B07740 standard; Protein; 787 AA.

AC B07740;

XX

XX

XX

DT 07-NOV-2000 (first entry)

XX

XX A snake venom protease (SVPH-1) polypeptide variant SVPH-1b.

DE

XX Snake venom protease; SVPH; SVPH-1; metalloproteinase-disintegrin;

KW chromosome 1; chromosome 4; immune system; splice variant.

XX

XX Homo sapiens.

OS

XX WO200043525-A2.

XX

XX 27-JUL-2000.

PD

XX

XX 21-JAN-2000; 2000WO-US01338.

PF

XX

XX 21-JAN-1999; 99US-0116670.

PR

XX 14-JUN-1999; 99US-0138682.

PR

XX 27-SEP-1999; 99US-0155798.

XX

XX (IMV ) IMMUNEX CORP.

PA

XX

XX Cerretti DP;

PI

XX WPI; 2000-482914/42.

DR

DR N-PSDB; A59305.

DR

XX WPI: 2000-482914/42.  
 DR N-PSDB; A59306.  
 XX  
 PT Snake venom protease (SVPH) nucleic acids, and polypeptides, used to  
 PT identify proteins having metalloproteinase-disintegrin activity, and  
 PT inhibitors of the proteins for use in therapeutics -  
 XX  
 PS Claim 13; Page 16; 105pp; English.  
 XX  
 CC The present sequence represents an alternatively spliced snake venom  
 CC protease-1 (SVPH-1) polypeptide, designated SVPH-1c. The SVPH  
 CC polypeptides are metalloproteinase-disintegrin protein family members.  
 CC The SVPH polynucleotides can be used as probes to identify nucleic acids  
 CC encoding proteins having metalloproteinase-disintegrin activity, to  
 CC identify human chromosome 1 or 4, to map genes on those chromosomes,  
 CC to identify genes associated with diseases, syndromes and conditions  
 CC associated with the chromosomes, and to study proteinases and their  
 CC activities on cell/cell interactions and the immune system. Sense or  
 CC antisense oligonucleotides of SVPH can be used to inhibit gene  
 CC expression of SVPH 1, 3, or 4. The SVPH polypeptides can be used to  
 CC study cell/cell and cell/matrix interactions involved in cellular  
 CC processes and in the immune system. The polypeptides may also be used to  
 CC screen for inhibitors of the polypeptide's activity, which are used in  
 CC therapeutics. The antibodies can be used in assays to detect the  
 CC presence of the polypeptides in vitro or in vivo, and to purify the  
 CC polypeptides by affinity chromatography.  
 XX  
 SQ Sequence 820 AA;

Query Match 11.3%; Score 80.5; DB 21; Length 820;  
 Best Local Similarity 22.3%; Pred. No. 0.74;  
 Matches 31; Conservative 28; Mismatches 47; Indels 33; Gaps 6;

QY 18 GKTPTHLK-----DGVCALYNEQDEAAVLEVPQHSLSLLHCRILEADPQTSITL-- 69  
 Db 60 gqkhihikvkkilfskhpvftydgailedpfvqnnocyh-gyvedgpesivslst 118  
 QY 70 -----YSMLQLQNFEMAMRCCLALDELHNVR-LCFOQSLEHL-----DEAFSDIVSG 118  
 Db 119 cfggfggilqin-----dfayeikplafsttfelhyvknseekqfstmrsg 165  
 QY 119 FIEHAAEVREYIAQLDESS 137  
 Db 166 fmqneitormefeidnst 184

RESULT 8  
 W44864  
 ID W44864 standard; Protein; 1105 AA.  
 XX  
 AC W44864;  
 XX  
 XX 28-AUG-1998 (first entry)  
 XX  
 DE Human TPC2 telomere length and telomerase regulatory protein.  
 XX  
 KW TPC2; telomere length; telomerase; human; cancer; gene therapy;  
 KW diagnosis; vaccine.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 182 /note= "encoded by TYT"  
 FT Misc-difference 211 /note= "encoded by NGG"  
 FT Misc-difference 217 /note= "encoded by GCT, C being uncertain"  
 FT Misc-difference 231 /note= "encoded by ATC, C being uncertain"  
 FT Misc-difference 381 /note= "encoded by ATC, C being uncertain"

FT Misc-difference 383 /note= "encoded by TTA, T being uncertain"  
 FT /note= "encoded by AYT"  
 FT Misc-difference 387 /note= "encoded by GGC, the second G being uncertain"  
 FT /note= "encoded by GAG, the second G being uncertain"  
 FT Misc-difference 396 /note= "encoded by GAG, the second G being uncertain"  
 FT /note= "encoded by GAG, the second G being uncertain"  
 FT Misc-difference 403 /note= "encoded by GAG, the second G being uncertain"  
 FT Misc-difference 404 /note= "encoded by GAG, the second G being uncertain"  
 FT Misc-difference 409 /note= "encoded by GAG, the second G being uncertain"  
 FT /note= "encoded by GAG, the second G being uncertain"  
 FT Misc-difference 410 /note= "encoded by GAG, the second G being uncertain"  
 FT /note= "encoded by GAG, the second G being uncertain"  
 FT Misc-difference 413 /note= "encoded by GAG, the second G being uncertain"  
 FT Misc-difference 416 /note= "encoded by GAG, the second G being uncertain"  
 FT /note= "encoded by GAG, the second G being uncertain"  
 FT Misc-difference 419 /note= "encoded by GAG, the second G being uncertain"  
 FT /note= "encoded by GAG, the second G being uncertain"  
 FT Misc-difference 421 /note= "encoded by GAG, the second G being uncertain"  
 FT /note= "encoded by GAG, the second G being uncertain"  
 FT Misc-difference 426 /note= "encoded by GAG, the second G being uncertain"  
 FT /note= "encoded by GAG, the second G being uncertain"  
 FT Misc-difference 434 /note= "encoded by GAG, the second G being uncertain"  
 FT /note= "encoded by GAG, the second G being uncertain"  
 FT Misc-difference 459 /note= "encoded by GAG, the second G being uncertain"  
 FT /note= "encoded by GAG, the second G being uncertain"  
 FT Misc-difference 460 /note= "encoded by GAG, the second G being uncertain"  
 FT /note= "encoded by GAG, the second G being uncertain"  
 FT Misc-difference 464 /note= "encoded by GAG, the second G being uncertain"  
 FT /note= "encoded by GAG, the second G being uncertain"  
 FT Misc-difference 469 /note= "encoded by GAG, the second G being uncertain"  
 FT /note= "encoded by GAG, the second G being uncertain"  
 FT Misc-difference 476 /note= "encoded by GAG, the second G being uncertain"  
 FT /note= "encoded by GAG, the second G being uncertain"  
 FT Misc-difference 477 /note= "encoded by GAG, the second G being uncertain"  
 FT /note= "encoded by GAG, the second G being uncertain"  
 FT Misc-difference 487 /note= "encoded by GAG, the second G being uncertain"  
 FT /note= "encoded by GAG, the second G being uncertain"  
 FT Misc-difference 488 /note= "encoded by GAG, the second G being uncertain"  
 FT /note= "encoded by GAG, the second G being uncertain"  
 FT Misc-difference 491 /note= "encoded by GAG, the second G being uncertain"  
 FT /note= "encoded by GAG, the second G being uncertain"  
 FT Misc-difference 510 /note= "encoded by GAG, the second G being uncertain"  
 FT /note= "encoded by GAG, the second G being uncertain"  
 FT Misc-difference 515 /note= "encoded by GAG, the second G being uncertain"  
 FT /note= "encoded by GAG, the second G being uncertain"  
 FT Misc-difference 517 /note= "encoded by GAG, the second G being uncertain"  
 FT /note= "encoded by GAG, the second G being uncertain"  
 FT Misc-difference 527 /note= "encoded by GAG, the second G being uncertain"  
 FT /note= "encoded by GAG, the second G being uncertain"  
 FT Misc-difference 529 /note= "encoded by GAG, the second G being uncertain"  
 FT /note= "encoded by GAG, the second G being uncertain"  
 FT Misc-difference 530 /note= "encoded by GAG, the second G being uncertain"  
 FT /note= "encoded by GAG, the second G being uncertain"  
 FT Misc-difference 551 /note= "encoded by GAG, the second G being uncertain"  
 FT /note= "encoded by GAG, the second G being uncertain"  
 FT Misc-difference 562 /note= "encoded by GAG, the second G being uncertain"  
 FT /note= "encoded by GAG, the second G being uncertain"  
 FT Misc-difference 563 /note= "encoded by GAG, the second G being uncertain"  
 FT /note= "encoded by GAG, the second G being uncertain"  
 FT Misc-difference 569 /note= "encoded by GAG, the second G being uncertain"  
 FT /note= "encoded by GAG, the second G being uncertain"  
 FT Misc-difference 573 /note= "encoded by GAG, the second G being uncertain"  
 FT /note= "encoded by GAG, the second G being uncertain"

FT Misc-difference 575 "encoded by YTC, C being uncertain"  
 FT Misc-difference 578 /note= "encoded by CCC, the third C being  
 FT /note= uncertain"  
 FT Misc-difference 586 "encoded by YTT"  
 FT Misc-difference 597 /note= "encoded by GCG, the first G being  
 FT /note= uncertain"  
 FT Misc-difference 599 /note= "encoded by CGC, G being uncertain"  
 FT Misc-difference 606 /note= "encoded by GAR, G being uncertain"  
 FT Misc-difference 608 /note= "encoded by GAG, G being uncertain"  
 FT Misc-difference 622 /note= "encoded by AMC"  
 FT Misc-difference 643 /note= "encoded by AAG, G being uncertain"  
 FT Misc-difference 645 /note= "encoded by RAC, C being uncertain"  
 FT Misc-difference 647 /note= "encoded by GAG, G being uncertain"  
 FT W09811204-A1.  
 PN 19-MAR-1998.  
 XX 13-SEP-1996; 96WO-US14679.  
 XX 13-SEP-1996; 96WO-US14679.  
 PR (GERO-) GERON CORP.  
 PA Adams RR, Andrews WH, Feng J, Villeponteau B;  
 PI WPI: 1998-207373/18.  
 XX N-PSDB; V19479.  
 DR Human TPC2, TPC3 and TR genes - regulate telomere length or modulate  
 PT telomerase activity  
 PS Disclosure; Fig 4A-G; 86pp; English.  
 XX This polypeptide comprises human TPC2, a protein that regulates  
 CC telomere length or modulates telomerase activity. Its amino acid  
 CC sequence was deduced from a cDNA clone (see V19479) obtained from a  
 CC human 293 cell library. TPC2 and TPC3 (see W44865) proteins can be  
 CC obtained by purification from natural sources, by in vitro  
 CC synthesis or by purification from recombinant host cells. They  
 CC have application in methods for reconstituting in vitro telomerase  
 CC or other enzymatic activities that maintain telomeres and regulate  
 CC telomere length. These methods have application in screens for  
 CC therapeutic agents, and for diagnostic tests. In addition,  
 CC peptides corresponding to TPC2 or TPC3 proteins can also be used to  
 CC regulate telomere length and telomerase activity in mammalian  
 CC cells. Immunogenic peptides and proteins of the invention can also  
 CC be used in therapeutic immunisation and vaccination procedures.  
 CC Antibodies that specifically bind to TPC2 or TPC3 proteins can  
 CC be used in screening, diagnosing and monitoring diseases and other  
 CC conditions, such as cancer, pregnancy or fertility.  
 XX Sequence 1105 AA;

Query Match 10.7%; Score 76.5; DB 19; Length 1105;  
 Best Local Similarity 20.0%; Pred. No. 3.4;  
 Matches 25; Conservative 31; Mismatches 46; Indels 23; Gaps 3;  
 QY 25 LKDGVCALYNEQDEAAVLEV-----POHSDSLLLHCRITTEADPQTSTLYSML 73  
 Db 691 ledkiralkenkqlesvlelhqrmegyrddpqhlekiaaygqkllqed-----lvhir 744

QY 74 LQNFEMAMRGCGWLALDELHNRLCFOQSLHLDSEAFSDIVSGFIEHAAEVREYIAQL 133  
 Db 745 aelsrestemenawneylklendveqlkqtqeqhrraf-----ffqeksqikqkdlwri 798  
 QY 134 DESSA 138  
 Db 799 edvta 803  
 RESULT 9  
 W73958  
 ID W73958 standard; Protein; 1105 AA.  
 AC W73958;  
 XX 29-APR-1999 (first entry)  
 DT Human TPC2 protein.  
 DE TPC2; TPC3; human; telomere length regulation; cancer; pregnancy;  
 KW fertility; diagnosis; therapy.  
 KW Homo sapiens.  
 OS Key Location/Qualifiers  
 FT Misc-difference 1..1105  
 FT /note= "X= unspecified amino acid"  
 XX US5858777-A.  
 XX 12-JAN-1999.  
 XX 13-SEP-1996; 96US-0710249.  
 XX 08-SEP-1995; 95US-0003492.  
 PR 05-JAN-1996; 96US-0583808.  
 PR 13-SEP-1996; 96US-0710249.  
 XX (GERO-) GERON CORP.  
 PA Adams RR, Andrews WH, Feng J, Villeponteau B;  
 PI WPI: 1999-152104/13.  
 DR N-PSDB; X01533.  
 XX DNA encoding proteins TPC2 and TPC3 - useful for regulating telomere  
 PT length or modulating telomerase activity  
 PS Claim 1; Column 47-52; 59pp; English.  
 XX This sequence is the human TPC2 protein, which is contained within  
 CC the recombinant mammalian host cell of the invention. The invention  
 CC provides methods and reagents for regulating telomere length and  
 CC modulating telomerase activity in mammalian cells as well as for  
 CC detecting, diagnosing, and treating related diseases and conditions such  
 CC as cancer, pregnancy, or fertility in humans and other mammals.  
 XX Sequence 1105 AA;

Query Match 10.7%; Score 76.5; DB 20; Length 1105;  
 Best Local Similarity 20.0%; Pred. No. 3.4;  
 Matches 25; Conservative 31; Mismatches 46; Indels 23; Gaps 3;

QY 25 LKDGVCALYNEQDEAAVLEV-----POHSDSLLLHCRITTEADPQTSTLYSML 73  
 Db 691 ledkiralkenkqlesvlelhqrmegyrddpqhlekiaaygqkllqed-----lvhir 744  
 QY 74 LQNFEMAMRGCGWLALDELHNRLCFOQSLHLDSEAFSDIVSGFIEHAAEVREYIAQL 133  
 Db 745 aelsrestemenawneylklendveqlkqtqeqhrraf-----ffqeksqikqkdlwri 798



QY 134 DESSA 138  
Db 799 edvta 803

## RESULT 10

W38542  
ID W38542 standard; Protein; 243 AA.  
XX W38542;  
AC  
XX  
DT 06-NOV-1998 (first entry)  
XX

DE Streptococcus pneumoniae protein of unknown function.

XX Streptococcus pneumoniae protein; genetic immunisation; antagonist;  
KW immunological response; inhibition; antibody production; inhibitor;  
KW T cell immune response; antimicrobial compound; bacterial adhesion;  
KW extracellular matrix protein; protein-mediated cell invasion; wound;  
KW pathogenesis.

XX Streptococcus pneumoniae.

XX WO9743303-A1.

XX 20-NOV-1997.

XX 14-MAY-1997; 97WO-US07950.

XX 14-MAY-1996; 96US-0017670.

XX (SMIK ) SMITHKLINE BEECHAM CORP.

XX (SMIK ) SMITHKLINE BEECHAM PLC.

XX Black MT, Hodgson JE, Knowles DJC, Nicholas RO;

XX Stodola RK;

XX WPI; 1998-008793/01.

XX N-PSDB; T98598.

XX Novel Streptococcus pneumoniae proteins and related DNA - useful for  
PT diagnosing anti-microbial agents for treatment of bacterial  
PT infections

XX Claim 12; Pages 320-321; 483pp; English.

XX This sequence represents a Streptococcus pneumoniae protein of  
CC unknown function, and is encoded by a DNA sequence of the invention.  
CC The DNA sequences were isolated from Streptococcus pneumoniae strain  
CC 0100993 (NCIMB 40794). The Streptococcus pneumoniae proteins of the  
CC invention can be used to identify compounds which interact with and  
CC inhibit or activate the activity of the proteins. Antagonists can be  
CC used to treat diseases caused by S. pneumoniae proteins, through genetic  
CC immunisation. They can also be used to induce an immunological response  
CC in a mammal by inoculation with the S. pneumoniae proteins or delivery  
CC of the encoding nucleic acids in a vector adequate to produce antibody  
CC and/or T cell immune responses to protect the animal from disease. The  
CC proteins can also be used to identify antimicrobial compounds which are  
CC capable of inhibiting their bioactivity. In particular the proteins of  
CC the invention can be used to prevent adhesion of bacteria to mammalian  
CC extracellular matrix proteins on in-dwelling devices or in wounds, to  
CC block protein-mediated mammalian cell invasion, and to block the normal  
CC progression of pathogenesis in infections initiated other than by the  
CC implantation of in-dwelling devices or other surgical techniques.

XX Sequence 243 AA;

Query Match 10.5%; Score 75; DB 19; Length 243;  
Best Local Similarity 23.8%; Pred. No. 0.62;  
Matches 29; Conservative 25; Mismatches 30; Indels 38; Gaps 7;

QY 20 KTIHLKDG--VCA-----LYNEQDEEAAVLEVPQHSDSLLLHC 56

Db 106 kkpissrdgmklcvetsttfdwvrqsekdyqdmliylkendfakigelte-knalamh- 163  
QY 57 RIIEADPQTSITLYSMLQLNFE--MAAMR-----GCWLALDELHNVL-CFOQSLEHL 107  
Db 164 ----attktaspafsyldasyeamafvrlrekeacyftmdagpnkvfcqekdehl 219  
QY 108 DE 109  
Db 220 se 221

## RESULT 11

Y85961

ID Y85961 standard; Protein; 317 AA.

XX Y85961;

XX 10-APR-2000 (first entry)

XX S. pneumoniae derived protein #170.

XX Treatment; prevention; disease; diagnosis; gene therapy; screening;  
KW bacterial; antimicrobial; antibiotic; pathogenesis; infection.

XX Streptococcus pneumoniae.

XX WO9806734-A1.

XX 19-FEB-1998.

XX 15-AUG-1997; 97WO-US14436.

XX 16-AUG-1996; 96US-0024022.

XX (SMIK ) SMITHKLINE BEECHAM CORP.

XX Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO;

XX Stodola RK;

XX WPI; 1998-159452/14.

XX N-PSDB; Z96312.

XX Streptococcus pneumoniae proteins and related DNA - useful for  
PT screening compounds for antibacterial activity

XX Claim 5; Page 456-457; 640pp; English.

XX This invention describes novel isolated Streptococcus pneumoniae  
CC polynucleotides (see Z96173-Z96494) and their encoded proteins (see  
CC Y85792-Y86182). The DNA, vectors and host cells described in the method  
CC of the invention are useful for the recombinant expression of the  
CC polypeptides. The polypeptides are useful for treatment or prevention of  
CC disease, or diagnosis of disease related to expression or activity of  
CC such a polypeptide. They can also be used to screen for compounds which  
CC interact with and inhibit or activate such a polypeptide. The  
CC polypeptides (or DNA encoding them, via gene therapy) are also useful  
CC for inducing an immunological response in a mammal. The antagonists are  
CC useful to inhibit such bacterial polypeptides. The polypeptides are  
CC particularly useful to identify antimicrobial compounds and antibiotics.  
CC They are also useful to determine their role in pathogenesis of  
CC infection, dysfunction and disease.

XX Sequence 317 AA;

Query Match 10.5%; Score 75; DB 19; Length 317;  
Best Local Similarity 23.8%; Pred. No. 0.9;  
Matches 29; Conservative 25; Mismatches 30; Indels 38; Gaps 7;

QY 20 KTIHLKDG--VCA-----LYNEQDEEAAVLEVPQHSDSLLLHC 56

Db 180 kkpissrdgmklcvetsttfdwvrqsekdyqdmliylkendfakigelte-knalamh- 237

QY 57 RIIADPQTSITLYSMLLQNFEE-MAAMR-----GCWLALDELHNVRL-CFQOQSLEHL 107  
 Db 238 ----attktaspafsyitdasyeamafvrqlrekgeacyftmdagpnvkvfcqekdlehl 293  
 QY 108 DE 109  
 Db 294 se 295

RESULT 12  
 B13516  
 ID B13516 standard; Protein; 317 AA.  
 XX  
 AC B13516;  
 XX  
 DT 02-NOV-2000 (first entry)  
 XX  
 DE Streptococcus pneumoniae mevalonate diphosphate decarboxylase.  
 XX  
 KW Mevalonate diphosphate decarboxylase; mvd; otitis media;  
 KW conjunctivitis; pneumonia; bacteraemia; meningitis; sinusitis;  
 KW pleural empyema; endocarditis; bacterial infection; Helicobacter pylori;  
 KW stomach cancer; ulcer; gastritis.  
 XX  
 OS Streptococcus pneumoniae.  
 XX  
 PN WO200044764-A1.  
 XX  
 PD 03-AUG-2000.  
 XX  
 PF 19-JAN-2000; 2000WO-US01132.  
 XX  
 PR 28-JAN-1999; 99US-0238477.  
 XX  
 PA (SMIK ) SMITHKLINE BEECHAM CORP.  
 XX  
 PI Wilding EI, Gwynn M, Iordanescu M;  
 XX  
 DR NPI; 2000-482967/42.  
 DR N-PSDB; A63347.  
 XX  
 PT New mvd (mevalonate diphosphate decarboxylase) polypeptide useful for  
 PT identifying its agonists and antagonists which may be used for  
 PT treatment and diagnosis of bacterial infections -  
 XX  
 PS Claim 1; Page 3; 38pp; English.  
 XX  
 CC The present sequence is the Streptococcus pneumoniae mevalonate  
 CC diphosphate decarboxylase (mvd) protein. The protein, gene, agonists and  
 CC antagonists can be used to diagnose and treat bacterial diseases,  
 CC particularly those caused by Streptococcus pneumoniae, such as otitis  
 CC media, conjunctivitis, pneumonia, bacteraemia, meningitis, sinusitis,  
 CC pleural empyema and endocarditis, and those caused by Helicobacter  
 CC pylori, including stomach cancer, ulcers and gastritis.  
 XX  
 SQ Sequence 317 AA;

Query Match 10.5%; Score 75; DB 21; Length 317;  
 Best Local Similarity 23.8%; Pred. No. 0.9;  
 Matches 29; Conservative 25; Mismatches 30; Indels 38; Gaps 7;

QY 20 KTIPIHLKDG--VCA-----LYNEQDEEAALVLEVPQHSLSLLHC 56  
 Db 180 kkpissrdgmklcvetsttdvvrqsekdyqdmliylkendfakigelte-knalamh- 237  
 QY 57 RIIADPQTSITLYSMLLQNFEE-MAAMR-----GCWLALDELHNVRL-CFQOQSLEHL 107  
 Db 238 ----attktaspafsyitdasyeamafvrqlrekgeacyftmdagpnvkvfcqekdlehl 293  
 QY 108 DE 109

Db 294 se 295

RESULT 13

B63634

ID B63634 standard; Protein; 466 AA.

XX

AC B63634;

XX

DT 26-MAR-2001 (first entry)

XX

DE Human gastric cancer associated antigen protein sequence SEQ ID NO:996.

XX

KW Human breast cancer; gastric cancer; prostate cancer; diagnosis;

KW cancer associated antigen; cytostatic; cancer vaccine.

XX

OS Homo sapiens.

XX

PN WO200073801-A2.

XX

PD 07-DEC-2000.

XX

PF 26-MAY-2000; 2000WO-US14749.

XX

PR 28-MAY-1999; 99US-0136526.

XX

PR 10-SEP-1999; 99US-0153454.

XX

PA (LUDW-) LUDWIG INST CANCER RES.

XX

PI Obata Y;

XX

DR WPI; 2001-025274/03.

XX

PT Nucleic acids encoding breast, gastric and prostate cancer associated  
 PT antigen precursors, useful for diagnosing and treating a condition  
 PT characterized by expression of an abnormal amount of a protein, e.g.  
 PT cancer -

XX

PS Example 1; Page 647-648; 799pp; English.

XX

CC F22422 to F22626, F22627 to F22773 and F22774 to F23014 represent  
 CC nucleotide sequences encoding human breast, gastric and prostate cancer  
 CC associated antigen precursors (CAAP) respectively. B63232 to B63467,  
 CC B63468 to B63721 and B63722 to B63970 represent human breast, gastric  
 CC and prostate CAAP protein sequence respectively. CAAPs have cytostatic  
 CC activity and can be used in the production of cancer vaccines. The human  
 CC CAAP proteins, peptides, nucleic acids or anti-CAAP antibodies are  
 CC useful for diagnosing and treating a condition characterised by  
 CC expression of an abnormal amount of a protein, e.g. cancer.

XX Sequence 466 AA;

SQ

Query Match 10.1%; Score 72.5; DB 22; Length 466;

Best Local Similarity 26.0%; Pred. No. 3.1;

Matches 32; Conservative 21; Mismatches 55; Indels 15; Gaps 5;

QY 24 HLKDGVCALYNEQDEEAALVLEVPQHSLSLLHCRIIADPQTSITLYSMLLQNFEE---- 79

Db 206 hekeaviclqnekdgkxemenimhsq----nceiklqksreivied-lkklhvndek 260

QY 80 MAAMRGCVLALDELHNVRLCFQOQSLEHLDEASFSDIVS----GFIEHAAEVREYIAQLDE 135

Db 261 kqliraelqslqshkeledtlqvrhidge--fekvmtdhvrsleelkknqgqlinqiqe 318

QY 136 SSA 138

Db 319 sha 321

RESULT 14

W82500

ID W82500 standard; Protein; 920 AA.

```

XX W82500;
XX
XX 01-FEB-1999 (first entry)
XX
XX Human OGT protein.
XX
XX OGT; O-linked GlcNAc transferase; uridine: transferase; human; tumour;
XX diphospho-N-acetylglucosamine; polypeptide beta-N-acetylglucosaminyl;
XX predilection; type II diabetes; hyperglycaemia; Alzheimer's disease;
XX metastasis; diagnosis; inhibitor; treatment; diabetes mellitus.
XX
XX Homo sapiens.
XX
XX W09844123-A2.
XX
XX 08-OCT-1998.
XX
XX 27-MAR-1998; 98WO-US06101.
XX
XX 31-MAR-1997; 97US-0042270.
XX
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
XX Hanover JA, Lubas W;
XX
XX WPI; 1998-557118/47.
XX
XX N-PSDB; V69301.
XX
XX Protein exhibiting O-linked GlcNAc transferase activity, OGT -
XX useful, e.g. to assess predilection to type II diabetes or
XX Alzheimer's or metastatic potential of tumours, and to identify
XX inhibitors
XX
XX Claim 7; Page 31-33; 56pp; English.
XX
XX This sequence represents a novel human O-linked GlcNAc transferase, OGT
XX protein (also known as uridine diphospho-N-acetylglucosamine:
XX polypeptide beta-N-acetylglucosaminyl transferase). This protein is
XX useful to assess predilection toward type II diabetes in patients
XX suspected of having hyperglycaemia that could evolve into this disease,
XX by assaying OGT activity in red blood cells. It can also be used to
XX assess predilection toward Alzheimer's disease, to assess the
XX metastatic potential of tumours and to diagnose a tumour with metastatic
XX potential. OGT can also be used to identify OGT inhibitors, especially in
XX high-throughput assays, useful, e.g. in the treatment of diabetes
XX mellitus, tumour-derived diseases and Alzheimer's disease.
XX
XX Sequence 920 AA;

Query Match          10.0%; Score 71.5; DB 19; Length 920;
Best Local Similarity 24.4%; Pred. No. 11;
Matches 30; Conservative 21; Mismatches 51; Indels 21; Gaps 4;

QY 24 HUKDGVYCALYNEQDEAAVLEV-----POHSDSL--LLHCRIRIADPOTSTLY 70
DB 165 hfpdaycnlanalkelgsvaeacdnyntalrlcpthadslnnlanikredgniteavrlly 224

QY 71 SMLLQNFEMAAARGCWLALDELHNVRLCFQOSLEHLEDEA-----SFSDIVSGFIEHAAE 125
DB 225 rkalevfepfaaahs---nlavslqggkqlqaleahmykeairisptiadaysmngntlke 281

QY 126 VRE 128
DB 282 mqd 284

RESULT 15
ID W48708
XX W48708 standard; Protein; 2183 AA.
XX
XX W48708;

```

```

XX 13-OCT-1998 (first entry)
XX
XX Measles virus Moraten vaccine L protein.
XX
XX L protein; attenuation; non-segmented; negative sense; vaccine; immunity;
XX single stranded RNA virus; Mononegavirales.
XX
XX Measles virus.
XX
XX W09813501-A2..
XX
XX 02-APR-1998.
XX
XX 19-SEP-1997; 97WO-US16718.
XX
XX 27-SEP-1996; 96US-0026823.
XX
XX (AMCY ) AMERICAN CYANAMID CO.
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
XX Murphy BR, Randolph VB, Sidhu MS, Tatem JM, Udem SA;
XX
XX WPI; 1998-230710/20.
XX
XX N-PSDB; W48708.
XX
XX Recombinantly-generated, attenuated, non-segmented, negative-sense,
XX single stranded RNA virus of order Mononegavirales - having
XX attenuating mutation in 3' genomic promoter region and RNA
XX polymerase gene, useful as vaccine to immunise against such virus
XX
XX Disclosure; Page 190-198; 426pp; English.
XX
XX This sequence represents the L protein from Measles virus Moraten
XX vaccine. This sequence is used in a method which involves the
XX isolation of recombinantly-generated, attenuated, non-segmented,
XX negative-sense, single stranded RNA virus of the order Mononegavirales
XX which have at least 1 attenuating mutation in the 3' genomic promoter
XX region and at least 1 attenuating mutation in the RNA polymerase gene.
XX This RNA virus can be used as a vaccine to immunise an individual against
XX such a virus.
XX
XX Sequence 2183 AA;

Query Match          9.9%; Score 70.5; DB 19; Length 2183;
Best Local Similarity 20.0%; Pred. No. 48;
Matches 38; Conservative 26; Mismatches 63; Indels 63; Gaps 7;

QY 4 SOQRVERFLOQYFSAGCK-----TPHLKDGVCALYNEQDEAAVLEVP 46
DB 158 ssqwfepflfwitvktcmrsviksgthchrtrhtpvtfgssvelliisrldvaliskes 217

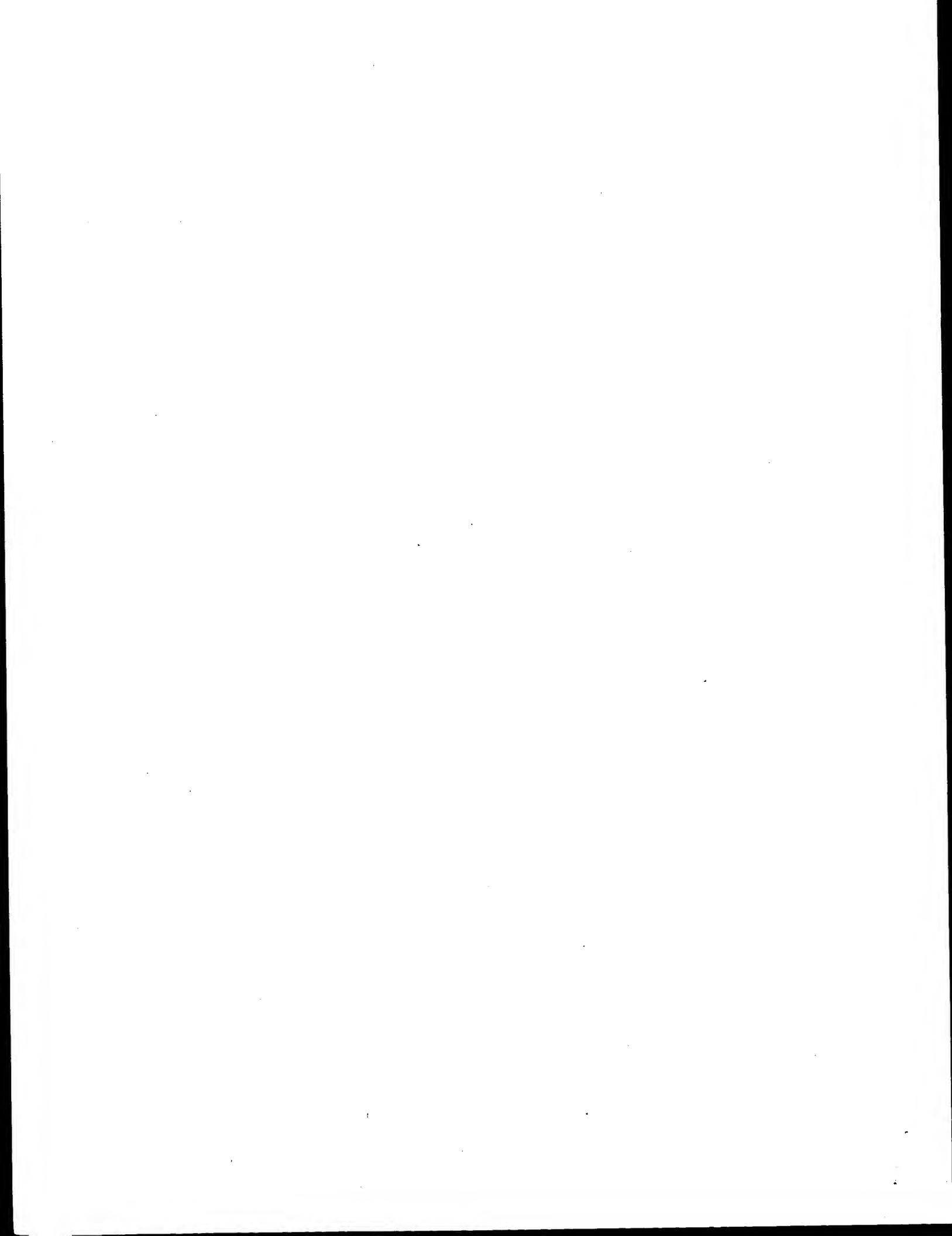
QY 47 OH-----SDSLHCRIRIAD--POTSITLYSMLLQNFEMAAARGCWLALD----- 91
DB 218 qhvvvylttfelvmvycdviegrlmtetamtidarytel---lgrvrymklidgffpalgn 274

QY 92 -----ELHNVRL--CFQOSLEHLEDEASFSDIVSGFIEHAAE 125
DB 275 ptyqivamleplslaylqlrditvelrgaflnhcfteihdvlqngfsd--egtyheite 332

QY 126 VREIAQLDE 135
DB 333 aldyifitdd 342

Search completed: June 5, 2001, 18:16:37
Job time: 203 sec

```



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2001, 18:14:14 ; Search time 54.28 Seconds  
(without alignments)  
2327.059 Million cell updates/sec

Title: US-09-596-784-2

Perfect score: 9448

Sequence: 1 MELKSLGTEHKAHVHTAAHN.....NPQVASALTDLKKEGLEMK 1838

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR\_67.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	9448	100.0	1838	2 T18448	pathogenicity factor
2	1824	19.3	1795	2 T30332	avirulence protein
3	275	2.9	2514	2 F81045	hemagglutinin/hemo
4	258.5	2.7	2248	2 A35938	profilaggrin - hum
5	255.5	2.7	3455	2 B82519	hemagglutinin-like
6	247	2.6	3442	2 B82589	hemagglutinin-like
7	245	2.6	2703	2 H81193	hemagglutinin/hemo
8	244.5	2.6	3259	1 A56539	giantin - human
9	243.5	2.6	3225	1 F52300	giantin - human
10	240.5	2.5	5327	2 T13564	microtubule-associ
11	237.5	2.5	4152	2 T31102	filamentous hemagg
12	232	2.5	1577	2 A35140	hemolysin A precu
13	229.5	2.4	2273	2 T09083	hemagglutinin/hemo
14	221	2.3	1608	2 A28182	hemolysin A - Serr
15	220.5	2.3	2253	2 T30336	nuclear/mitotic ap
16	219.5	2.3	2541	2 S11661	talin - mouse
17	218.5	2.3	4957	2 T03455	ALR protein - huma
18	218.5	2.3	5262	2 T03454	ALR protein - huma
19	217.5	2.3	1957	2 T38077	hypothetical colle
20	217	2.3	2712	2 T05113	hypothetical prote
21	214	2.3	3282	2 E82750	hemagglutinin-like
22	212.5	2.2	5627	2 C83339	hypothetical prote
23	210	2.2	2059	2 D82671	surface protein XF
24	209.5	2.2	1510	2 T33100	hypothetical prote
25	209.5	2.2	4319	2 T33105	hypothetical prote
26	207.5	2.2	3591	1 S21010	filamentous hemagg
27	205	2.2	1684	2 JW0057	gravin - human
28	204.5	2.2	4558	2 C82199	RTX toxin RtxA VC1
29	204	2.2	2845	2 I49505	adenomatous polyo

## ALIGNMENTS

## RESULT 1

T18448

pathogenicity factor DspA - Erwinia amylovora

C:Species: Erwinia amylovora

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 20-Jun-2000

C:Accession: T18448

R:Gaudriault, S.

submitted to the EMBL Data Library, May 1998

A:Reference number: Z18936

A:Accession: T18448

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1838 <GAU>

A:Cross-references: EMBL:Y13831; PIDN:CAA74156.1

C:Genetics:

A:Note: dspA

C:Function:

A:Description: involved in pathogenicity

Query Match

Best Local Similarity 100.0%; Score 9448; DB 2; Length 1838;

Matches 1838; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

30 203.5 2.2 1995 2 G81044  
31 203.5 2.2 3187 2 JC5837  
32 202 2.1 1975 2 B81192  
33 200 2.1 1467 2 A75564  
34 199 2.1 2232 2 T34434  
35 198.5 2.1 835 1 I57441  
36 197.5 2.1 2015 2 B81989  
37 196.5 2.1 1902 1 B4858  
38 195.5 2.1 4377 2 A55575  
39 194 2.1 2954 2 T14156  
40 193.5 2.0 1962 2 A32634  
41 192.5 2.0 2938 2 T30249  
42 192 2.0 1217 2 S2714  
43 192 2.0 2346 2 T13829  
44 190.5 2.0 1536 2 A43855  
45 189 2.0 1902 2 S06997

hemagglutinin/hemo  
364K Golgi complex  
hemagglutinin/hemo  
conserved hypothet  
hypothetical prote  
involucrin - orang  
hypothetical prote  
lactocepin (EC 3.4  
ankyrin 3, long sp  
kinesin-related pr  
lactocepin (EC 3.4  
cell proliferation  
sericin1B - silkw  
Tpr homolog - frui  
high-molecular-wei  
lactocepin (EC 3.4

QY 421 PGEAHSLLTGIWHPAGAAPQGESIRLHDDKTHILHPELGVWQSDAKDTHSOLROAD 480  
 DB 421 PGEAHSLLTGIWHPAGAAPQGESIRLHDDKTHILHPELGVWQSDAKDTHSOLROAD 480  
 QY 481 GKLYALKDNTLQNLSDNKSSEKLVYDKISYSDVQGOVALTTPGHRHKSIMPSLDAS 540  
 DB 481 GKLYALKDNTLQNLSDNKSSEKLVYDKISYSDVQGOVALTTPGHRHKSIMPSLDAS 540  
 QY 541 PESHISLSLHFAHAGHLLHGKSELEAQSAIASHGRLVADSEGLFSAAIIPKQGDGNEL 600  
 DB 541 PESHISLSLHFAHAGHLLHGKSELEAQSAIASHGRLVADSEGLFSAAIIPKQGDGNEL 600  
 QY 601 KMKAMPOHALDEHPGHQHDSGFFHDDHGHQNLVKNFNROQHACPLGNDHQFHPGWNLT 660  
 DB 601 KMKAMPOHALDEHPGHQHDSGFFHDDHGHQNLVKNFNROQHACPLGNDHQFHPGWNLT 660  
 QY 661 DALVIDNOLGLHHTNPPEHETLDMGHGLSLALQEGKLYHFDQLTGKWTGAESDCKQLKKG 720  
 DB 661 DALVIDNOLGLHHTNPPEHETLDMGHGLSLALQEGKLYHFDQLTGKWTGAESDCKQLKKG 720  
 QY 721 LDGAAYLLKDGVEKRLNINQSTSSIKHGTENVFSLPHVRNKPPEGDALQGLNKDDKAQAM 780  
 DB 721 LDGAAYLLKDGVEKRLNINQSTSSIKHGTENVFSLPHVRNKPPEGDALQGLNKDDKAQAM 780  
 QY 781 AVIGVNYLALTEKGDITRSFOIKPCTOOLERPACTLSREGTSGELKDIHVDHKQNLALYALT 840  
 DB 781 AVIGVNYLALTEKGDITRSFOIKPCTOOLERPACTLSREGTSGELKDIHVDHKQNLALYALT 840  
 QY 841 HEGEVFHPREAWONGAESSEWHKLLALPOSESKLSLSDMSHEHKPIATFEDGSQHOLKAG 900  
 DB 841 HEGEVFHPREAWONGAESSEWHKLLALPOSESKLSLSDMSHEHKPIATFEDGSQHOLKAG 900  
 QY 901 GWHAYAPERGLAVGTSGSTVFNRLMQGVKVPISGSLTVKLSAQOTGGMTGAEGRKV 960  
 DB 901 GWHAYAPERGLAVGTSGSTVFNRLMQGVKVPISGSLTVKLSAQOTGGMTGAEGRKV 960  
 QY 961 SSKTSERTRAVAFNPTMTFPRPKNAAYATQHWGREGKLPYEMOGALIKQLDAHNVR 1020  
 DB 961 SSKTSERTRAVAFNPTMTFPRPKNAAYATQHWGREGKLPYEMOGALIKQLDAHNVR 1020  
 QY 1021 HNAPOPDLQSKLETLDLGHEGALLNDMKRFRDELEQSAATRSVTVLGGHGVLSKNGEIN 1080  
 DB 1021 HNAPOPDLQSKLETLDLGHEGALLNDMKRFRDELEQSAATRSVTVLGGHGVLSKNGEIN 1080  
 QY 1081 SEFKPSCKALVQSFVNARSQDLSKSLQAAVHATPPSAESKLSQMLGHFVSAGVDMSHQ 1140  
 DB 1081 SEFKPSCKALVQSFVNARSQDLSKSLQAAVHATPPSAESKLSQMLGHFVSAGVDMSHQ 1140  
 QY 1141 KGEIPLGRQDPNDKTALTKSRLIILDTVTIGELHELADKAKLVSDHKKPADQIKQLRQOF 1200  
 DB 1141 KGEIPLGRQDPNDKTALTKSRLIILDTVTIGELHELADKAKLVSDHKKPADQIKQLRQOF 1200  
 QY 1201 DTLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKEHGVNLTTRTVLESQ 1260  
 DB 1201 DTLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKEHGVNLTTRTVLESQ 1260  
 QY 1261 SAEALAKLKNTLLSLDGSBMSFSRSYGGGVSTVFVPTLSKKVPVPVPGAGITLDRAYN 1320  
 DB 1261 SAEALAKLKNTLLSLDGSBMSFSRSYGGGVSTVFVPTLSKKVPVPVPGAGITLDRAYN 1320  
 QY 1321 LSFRTSGGLNYSFGRDGVSGNIMVATGHDVMPYMTGKTSAGNASDWLSAKHKISPD 1380  
 DB 1321 LSFRTSGGLNYSFGRDGVSGNIMVATGHDVMPYMTGKTSAGNASDWLSAKHKISPD 1380  
 QY 1381 RTGAASVGTLOGLTQNSLKFKLTDELPGFIHGLTHTLTPAELLQKGIIEHQMKQGSKLT 1440  
 DB 1381 RTGAASVGTLOGLTQNSLKFKLTDELPGFIHGLTHTLTPAELLQKGIIEHQMKQGSKLT 1440  
 QY 1441 FSVDTSANLDRAGINLINEDGSKPNGVTVARSAGLSASANLAAGSRERSTTSQGFSTTS 1500  
 DB 1441 FSVDTSANLDRAGINLINEDGSKPNGVTVARSAGLSASANLAAGSRERSTTSQGFSTTS 1500  
 QY 1501 ASNNRPTFLNGVAGANLTAALGVASHSTHEGKPVGIFPAFTSTNVSAALADNRTSOSI 1560

DB 1501 ASNNRPTFLNGVAGANLTAALGVASHSTHEGKPVGIFPAFTSTNVSAALADNRTSOSI 1560  
 QY 1561 SLELKRAEPTVNDISELTSTLGKHKFDSATTKMLAALKELDDAKPAEQHLILQOHSFAK 1620  
 DB 1561 SLELKRAEPTVNDISELTSTLGKHKFDSATTKMLAALKELDDAKPAEQHLILQOHSFAK 1620  
 QY 1621 DVVGDERYEAVERNKLKLVIRQQAADSHSMELGSASHSTTYNNLSRINNDGIVELLHKHFD 1680  
 DB 1621 DVVGDERYEAVERNKLKLVIRQQAADSHSMELGSASHSTTYNNLSRINNDGIVELLHKHFD 1680  
 QY 1681 AALPASSAKRLGEMMNDPALKDIIRKOLQSTPFSSASVSMELKDLREQTEKAILDGKVG 1740  
 DB 1681 AALPASSAKRLGEMMNDPALKDIIRKOLQSTPFSSASVSMELKDLREQTEKAILDGKVG 1740  
 QY 1741 REEVGVLFQDRNLRVKSVSQSKSEGFNTPALLIGTNSAAMERNIGTINFKYK 1800  
 DB 1741 REEVGVLFQDRNLRVKSVSQSKSEGFNTPALLIGTNSAAMERNIGTINFKYK 1800  
 QY 1801 QDQNTPRRTLEGGIAQANPQVASALTDLKKEGLEMS 1838  
 DB 1801 QDQNTPRRTLEGGIAQANPQVASALTDLKKEGLEMS 1838

RESULT 2  
 T30332  
 avirulence protein - Pseudomonas syringae  
 C:Species: Pseudomonas syringae  
 C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999  
 C:Accession: T30332  
 R:Bogdanove, A.J.; Kim, J.F.; Wei, Z.; Kolchinsky, P.; Charkowski, A.O.; Conlin, A.K.  
 P:U.S.A. 95, 1325-1330, 1998  
 A:Title: Homology and functional similarity of an hrp-linked pathogenicity locus, dsp  
 A:Reference number: Z20825; MUID:98115919  
 A:Accession: T30332  
 A:Status: Preliminary; translated from GH/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1795 <BOG>  
 A:Cross-references: EMBL:U97505; NID:g2978502; PID:g2978503; PIDN:AAC06134.1  
 C:Genetics:  
 A:Gene: avrE

Query Match 19.3%; Score 1824; DB 2; Length 1795;  
 Best Local Similarity 29.1%; Pred. No. 5.8e-83;  
 Matches 554; Conservative 358; Mismatches 722; Indels 272; Gaps 71;  
 QY 34 SSSPONAASLAAGKNGKMPRIHQPTAADGISAHQKQKFSRLGCLGTTKFSRSAP 93  
 DB 61 SKAPQQAAT-PPTAKN-----VKTTPPASNVATPRNKARES-----GFSNSSP 103  
 QY 94 Q-----GQPTTHSKGATLRDLLARDGDTQHEAAAPDAARLTRSGVKRRND 142  
 DB 104 QNTHRAPKWLNRHPNOASSGAQTHEI-----HPEAAP-----RKNLR 142  
 QY 143 DMAGRPVWKGSGEDKVTPTQQRHQLNFGOMQRTMLSKMAHPASANAGDRLQHSPPHIP 202  
 DB 143 VRFDLPQ-----DRLESPSYLD 160  
 QY 203 GSHHEIKREPVGSTSKATTADHADVIEIAQEDDDSEFQQLHOORLARERNPPPOP-----PK 258  
 DB 161 SDNPTDEAVANATRF-----RSPDHLQSGDGTTRISMLATDPDPQSSSGSK 209  
 QY 259 LGVAT-PLSARFQPKLTAVAESVLEGTDTTQSPKRPQSMKSGAGVTPPL-AVTLD-KGK 315  
 DB 210 IGDSGPIPPR-----EPLWRSNGRFEKDEKLVRNS-----EPQGSITLDKAGK 256  
 QY 316 LQAPDNPALNTLLKQTLGDKDTQHYLAHASSDGSQHLHLDNKGHLFDIKSTATSYVL 375  
 DB 257 PDFSTFNTPLGLAPLDSILATPKQYLAHQKDGKGVGHQLOLANGHFLHLAQDSSILAVI 316  
 QY 376 HNSHPG-BIKGLAQAGTGSVSDGKSGKISLGSGTQSHNKTMLSQFGEAHRSLITGIWQ 434  
 DB 376 HNSHPG-BIKGLAQAGTGSVSDGKSGKISLGSGTQSHNKTMLSQFGEAHRSLITGIWQ 434

Db 317 RSSNEALLIEGKKPPA---VKMEREDGNIHI--DTASGRKTO-ELPGKAHIAHITNV-- 367  
 QY 435 HPAGARPOGESIRLHDDKIHILHPELVWQSAD--KDT-HSOLSRQADGKLYALKDNRT 491  
 Db 368 ----LLSHDGERMRVHEDRLYFDPITSTRMKIPEGLEDYAFNSLSTGGGCVYAKSDAV 423  
 QY 492 LONLSDNKSSEKLVDKIKSYSDVDRGOVALLTDPGRHKWMSIMPSLDASPEHISLSLHF 551  
 Db 424 V-DLSPENPHVEVEDIQSFSVAPDNRAALLS---OKTQOAILLT---VIGGLT 474  
 QY 552 ADAHOGLLHGKSELEAQSVAISHGRILVADSEGLFSAATPKOCGDGNEMLKAMPOHALD 611  
 Db 475 PKKTGLELDGKGAAAVCLSGDKLFADTQGRLYSAD-RSAFEGDDPKLKLMPQANF 533  
 QY 612 E----HFHGDHQISGFFHDDHGOALNALVKNFRQOQACPLG-NDHFQHPGWNITDALVID 666  
 Db 534 QLEGVPLGGHNRVFTFINGDDGGVHALIKNRQGETHSHALDEOSSKLOSQWNLTNALVLN 593  
 QY 667 NOLGLHHTNPEP---HEILDGMHGLSLALQEGKLHVFDQLTGKWTCAE-SDCKQLKKGLED 722  
 Db 594 NNRLG--TWPPPTAADRLNLDRLAGLVGLSEGRIQRDATPECKWDAGTKIDRLQRGAD 651  
 QY 723 GAAYLLKQGEVKRLNINOSTSIKIGTENVFLPHVRNKPPEPDALQGLNKDDKA-QAMA 781  
 Db 652 SNAYVLKGGKLHALKIAAEPHMAFDNRTALAQATARSTKVEMGKETGL--DDRVIKAPA 709  
 QY 782 VIGVNYKYLALTEKDIRSFOIKPGTOOLERPQAQTLREGISGELKDIHVHDKONLYALTH 841  
 Db 710 MYSNKRFFVALDD-----QNKLTASHDKHPVTLDIPLEGDDIKSLSDKHNHALTS 762  
 QY 842 EGEVPHOPREAWQN---GAE-SSSHKHALPOSEKSLSDMSHEHKPIATFEDGSOH-- 895  
 Db 763 TGGLYCLPKEAWQSTKGLDQLRARPTVALPGQ-PVKALFTNDDNVLSAQIEDAEGKL 821  
 QY 896 -OLKAGGHAYAAAPRGPLAVCTSSQTVFNRLMGVKGKVIPIGSLTVKLKSAQTGCMGTG 954  
 Db 822 MOLKAGOMORE---EQRP--VEENGLNDVHSRITGSNTWRIPKGTGLTIRMVNTFGRSG 876  
 QY 955 AGRKVSXKFSERIRAYAFNPTMTSPRIKNAAYATQHWQCGREGKLPLEYMOGALIKOL 1014  
 Db 877 VEKSKAST-SEFIRANIKYKNAETPRWKNVGDHIOHRYQGRGLGKVEYTESMLFKQL 935  
 QY 1015 DAHNVNRNAPQ-----DLQSKLETLD--LGEHGAELNLMKRRFDEBQSATRSVTVIG 1067  
 Db 936 EL---IHESGGPPARGQDOLKARITALEAKLPGQATLVKELETLDLENHNSVTALMSG 993  
 QY 1068 -----OHQGVLSKNGEINSEFKPS-----GKALVQ-----SENVNRSQDLSKSLQ 1110  
 Db 994 QSYCKAKNLAKQDGIILNHQEL---AKPSVRMQFGKKLADLCTKLNFKSSGHDLVKELOD 1050  
 QY 1111 AVHATPPSAESKLOSLMGHFVSAGVDMSHQKEIPLGRQRPNDKRTALTAKSLIILDTVTI 1170  
 Db 1051 ALTQVAPSAENPTKLLGLTKHQGLKLSHQKADIPLGQRRDAASEDHGLSKARLADIVTL 1110  
 QY 1171 GELHELADKALVSDHDPADQIQLROQFDTLREKRYESNPVKHYTDMGTFHNALEAN 1230  
 Db 1111 KSLGALLDQV---EQLPPQSDIEPLQKKLATLDVTVYGENPVKYVTDWGMFTDNKALESG 1166  
 QY 1231 YDAVKAFINAFKKEHGHVNLTRTVLESQGSAAELAKLKNLTLLSDSG-ESMSFSRSKSG 1289  
 Db 1167 YESVKTEFLSKKADHADVSNVNRATGSKQOAEALAGFKSKMLKQLEHGDDEVLGRQSGV 1226  
 QY 1290 GVSTVFVPTLSKKVPVPIPGAGITLDRAYNLFSRSTSGGLN---VSFRGDGGVSGNIMV 1346  
 Db 1227 NLTPFTILADK--ATGLWPTAGATGNRYILNAERCEGGVTLYLISEGA-GNVSGGF-- 1281  
 QY 1347 ATGHDVMPYM-----TCKKTSAGNASDWLSAKHKLSPDLRTGAASVGLQGTQLNSLKF 1401  
 Db 1282 GAGKDYWPGFDDANNPARSDVGN-----NRTLTPNPLRGVDVDTATVAASQRAQVYFN 1334  
 QY 1402 LTEDPGFIHGLTHGLTTPAELLQKGIHOMKQSKLTFSSVDTSANLIDLRAGINLNE- 1460  
 Db 1335 VPDEDIDAFVDDLFEGLNPLQVLKKAVDHESYEAERFNFDLTAGGTADIRAGINLITE 1394

QY 1461 --GSKPN--GVTIARYSAGLSA--SANLAAGSRERSTTSGQFGSTTSASNNRPTFLNGVGA 1514  
 Db 1395 DPNADPNSDSFSFSAVVRGGFAANITVNLMTYTDYSLTKNDKTELKGGKRRPFLNNVTA 1454  
 QY 1515 GANLTAALGVCAHSSTHECKPV---GTFPA--FTSTNVSAAL--ALDNRTSQSISLEKRA 1567  
 Db 1455 GGQURAQGGSHATP-TGTPASAPGPTPASQTAANLGGALNFSVENRTVKRIKFRVNA 1513  
 QY 1568 EPTVTSNDISELTSLGKHKFKDSATTKMLAALKELDDA-----KPAE-----QLHILQOHFS 1618  
 Db 1514 KPITTEGUSKLSKGLGEAFLODNTTAKLAELADPLNARYTKGRKPDVEVIOAQDGLBELFA 1573  
 QY 1619 ----AKDVGVDEREAVRNKLKLVIROQA-ADSHSMELGSAHSHTTYNNLSRINNDGIV- 1672  
 Db 1574 DIPPPKO--NDQYKALRDLKRAAVEHRASANKHSV-MDNAREFETSKTSLGSLSESILT 1630  
 QY 1673 ELLJHGFDAALPASSAKRLEGEMNDNDPALKDIIKQLOSTPFSASVSMELKDLREOTEK 1732  
 Db 1631 KIMSVRDASAP-GNATRVAEFMRQDPKRLAMLKEMEGSIGTLARVRLEPKDSLVDKIDE 1689  
 QY 1733 AILDGKVGREEVGVLFDQRNRLRVKSVSQSVKSEGTTPALLIGTNSAAMSERNI 1792  
 Db 1690 GSLNGTWTQSDLSMLEDNRNEMRIKRLVVFHTATQAEFTSTPLVSYSGANVSVTKTL 1749  
 QY 1793 GTINFKYQDQONTPRFTLEGGIAOANPOVASALTDLKEGLEMK 1838  
 Db 1750 GRINFEVYADQDKPIGYTFDGLSRPSASLKEAAGDLKKEGFELKS 1795

## RESULT 3

F81045

hemagglutinin/hemolysin-related protein NMB1768 [imported] - Neisseria meningitidis  
 C:Species: Neisseria meningitidis  
 C>Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 31-Dec-2000  
 C:Accession: F81045

R:Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, R.; Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.; Ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masiugnani, V.; Pizza, M. Science 287, 1809-1815, 2000

A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.  
 A:Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58.  
 A:Reference number: A81000; MUID:2017555

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-2514 &lt;DET&gt;

A:Cross-references: GB:AE002098; GB:AE002526; GB:AE002098; NID:g7272015; PIDN:AAF42109.1; PID:g7272015

A:Experimental source: serogroup B, strain NMD58

C:Genetics:

A:Gene: NMB1768

Query Match 2.9%; Score 275; DB 2; Length 2514;

Best Local Similarity 18.8%; Pred. No. 3e-05;

Matches 389; Conservative 287; Mismatches 803; Indels 586; Gaps 91;

QY 7 GTEH-KAAVHTAAHNPV-----GHCVLQOQSSS-----SSPQNAAS 43  
 Db 37 GSAHVSVPPGTTTHAPVCRNIFSLGSLCLAVGTANTAFADGIIADKAAPKTOAT 96  
 QY 44 LAAGEKNGRKMPRIHOPSTAADGISAHQKKSFLRGCLGCTKKFSRSPGPOGP----- 98  
 Db 97 ILQTNG---IPOVNIOTPTTSAGVSVNOYQAFQVGNRGAILNN--SRNWTOTOLGWNTOG 151  
 QY 99 -----TTHSK-----GATRLDLARDDDGFTQHEAADAARLTR 132  
 Db 152 NPWLARGEARVVVQNINSSHSQMGVIEVGGRAEVVIANPAGIANGGGINASRATL 211  
 QY 133 SGGVKERNMDDMAGRPMVK-----GSGED-----KVPTQQRHOLNFGQNRQTMLS 180  
 Db 212 TTGQPOYQAGDLGFGFIROGNVVIAGHGLDARDTDFTRILSYHSKIDAPVWGVQVVRVAG 271







QY	1759	V-----SVSQSVSKSEGFN 1772	Db	1684	EKYR-----DAPRTTQDILNPDAGAPALIQTG-----GAMMIQTDLNRHYADLLAGDG 1733
Db	2206	AFGKDHPRYATYINKDPGLCGHSSDISKQLGFS 2239	QY	452	DKIHILPELGMVOSADKDTSHSQRADGKLYALKDNRTLQNLSDNKSSEKLVDKIKSY 511
			Db	1734	QTVGLPPHPTKEKSDDEHY-----KRVLLIDNRALQ-LS-----RTDTFHNH 1776
RESULT	5		QY	512	SVDQGVQA-----ILTDTPGRHKMSIMPDLADSPESHISLSLHFADAHQGLLHG 561
B82519		hemagglutinin-like secreted protein XF2775 [imported] - Xylella fastidiosa (strain 9a5c)	Db	1777	STTYRGKVEPWSNESRTPPTQIGRITSGGHQHIAAOTLANNVTDTSHAPEPIQHLTYN 1836
C:Date:	18-Aug-2000	#sequence_revision 20-Aug-2000 #text_change 20-Aug-2000	QY	562	KSELEAQVAISHGLRVVADSEGLFSAAPKOGDGNELKMKAMPOHAIHDEHGHCHQHS 621
C:Species:	Xylella fastidiosa		Db	1837	PS---TQTLVVNGVITVTDTSPLSHVTSADNGPSAGDELTYIPQDSITTP-----1885
C:Accession:	B82519		QY	622	GFHDDHCOLNALVK-----NN-----FRQOHACPLGNDHOFHPGWN 658
C:Anonymous:	The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen		Db	1886	-----NAPIRDPAAPPATVPTGPLPLPNNSLFTLHPDAATLITTDPRFLGRP 1935
C:Nature	406, 151-157, 2000		QY	659	LTDALVIDNOLGLHHTNPPEHILDMGHGLSLALOEGLKHYFDQLT--KGWTGAESCKQ 716
A:Title:	The genome sequence of the plant pathogen Xylella fastidiosa.		Db	1936	YTSADTQLHALGDHT---LHKRLGDGYEQRLIRE-----QLAOLTGRRRLDGYTDDQO 1988
A:Reference number:	A82515; MUID:20365717		QY	717	LKGLDGAAYLLKDGVEK-----RLNINQSTSSI-----KHGTENVFSLPHVN 760
A:Note:	for a complete list of authors see reference number A59328 below		Db	1989	YRALLDAGVTVAQHQHLPGLIADSLAQALTSDIVMLVOODVOLPDGTTTRALVPLRYL 2048
A:Accession:	B82519		QY	761	KPEPDALOGKANKDDKAQAMAVIGNVKYLALTEKGDIRSFOIKPGTQOLRPAOTLSREG 820
A:Status:	preliminary		Db	2049	RPTGD-----LPDGLALLAAASTINAH-TFTNTGTIDARHL-----IDINAHMTDQOQ 2097
A:Molecule type:	DNA		QY	821	-----ISGELKDIHVDHKNLYALTH-EGEVFHPQ--REAWQNGAESSSHWKL- 865
A:Residues:	1-3455 <SIM>		Db	2098	GRLTADAIHHTGDTFTLGGQFKARGYLKVHAGGNFLASSTLRDATQTGTRHSHVTELD 2157
A:Cross-references:	GB:AE004082; GB:AE003849; NID:99108003; PIDN:AAF85560.1; GSPDB:GN001		QY	866	-----ALPQSESKLSDMSHEKPIATFEDGSOHOLKAGGHAYAAAPERGPLAYGT- 917
A:Experimental source:	strain 9a5c		Db	2158	QOAGFTVTGPGAYLGL-STDOAMTHNGVAINTGTD-----GYTSLNA--TGPLHLCTL 2208
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A			QY	918	-----SGQTVFNRLMQGVKVPKPGSLTVKLSA-----947
R:Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H			Db	2209	TTHRSDDTQMDPRNSRHSRIDTEYGTSTIG-NGDIQNLNSQDINLRAATLHSTOCTITAL 2267
as-Neto, E.; Docena, C.; El-Dorriy, H.; Facincani, A.P.; Ferreira, A.J.S.			QY	948	QTGCMT-----GAEGKVKSKSERIRAYAFNPTMTSPRIKNAAY 988
submitted to GenBank, June 2000			Db	2268	ATGNVTITHGDTTQYTSQDSHTKRGSLNLSRTTTHADQOQTAIGSTLSADKVPVKGN 2327
A:Authors:	Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franco, S.C.; Franco, M.C.; Frohm		QY	989	ATOHGWOGREGLKPELQEMOGALIKOLDANVR-----HNAPODLOSKLETLDLGEHGAEL 1044
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig			Db	2328	ITVTGSHVSDAGTYMOAEHDLTLQAATHTTQSTYSHHTKORGL-----IRNGGASL 2379
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig			QY	1045	LNDMKFRDELBOSATRSVTVLGOHOG-----VLKSNGEINSSEKFPSP 1087
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E			Db	2380	TLGNQSQRTSTTTATTTTGLICATGNVTLLAGGHVQOIGSDVLSPHGDDIDIAKKVD 2439
A:Authors:	Martins, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A		QY	1088	GKALVQSFNVNRSQDL-----SKSQQAQVHATPPSAESKLSMLGH 1129
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak			Db	2440	---IIQAHHTSQTHATROSGLTVALSTPLLAGAQTAAQOMQHAARSGDPRLOALAGL 2496
A:Authors:	da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir		QY	1130	FVSAGVMSHQKGEIPICRQDPNDKATLTKSRLLDVTVTIGELHELADKAKLVSDHKPD 1189
M.; Tshukako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z			Db	2497	TTALGA-----KNTIDAVRO-DPRALGGLNAS-----LTVG-----2526
A:Reference number:	A59328		QY	1190	ADQIKLQKQQPDTLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKEHHGVN 1249
A:Contents:	annotation		Db	2527	-----RSHDSTTTTTTSAAGSNVTAGNVH-----ISATGDGTASTLTIOGSDVRG--2574
C:Genetics:			QY	1250	LTTTFTVLESQSAELAKKLKNTLLSDGESMSFSRSYG-----GGVSTVFVPTLS 1300
C:Gene:	XF2775		Db	2575	-DTMTYLKADGDIAL-AAQNTV-----TNQRNRSRAGVGVANVLGSGGTSAGLTAHAS 2628
			QY	1301	KKVPVPVPIGAGITLDRAYNLFSFRTSGG--LNVSFGRD-----GGVSGNI 1344

Query Match	2.7%	Score	255.5;	DB	2;	Length	3455;
Best Local Similarity	18.6%;	Pred. No.	0.00048;				
Matches	412;	Conservative	285;	Mismatches	825;	Indels	695;
						Gaps	101;
QY	3	LKSLCTEHAHVTAHNPVCHGVALOQSSSSSPONAAASLAEGKNRGMKPRITHQPS 62					
Db	1374	IDNLGTGRLYGDHIAL-----HAQTLNRDETSQHTHAATIAARQLDIDGADTLRNTAN 1428					
QY	63	A---ADGISAHQKKSFLRGCL---GTTKF--SRAPQOQPGTTHSKGATLRDLLARD 114					
Db	1429	AWILSDGDAI-----GATLONALHATGTATLLDNRSATIDITGTLNITTTTLNNI--RD 1481					
QY	115	DCETQHEAAPPD---ARLTSRGVKKRRNMDMAGRPVKGSGEDKVPQCKRHLNMF 171					
Db	1482	N---VHTAHAPDVVTEARMEQPHW--RNQPN-----GSGNFRFTSNYDAHDI---1525					
QY	172	GOMROTMLSKMAHPASANAGDLQSPHPHPSHHEIKEEPPVGSTK-----A 219					
Db	1526	-----YYLNPADIKDD-----PYITPDGQOIHRAIVRLTPQTSAYFARGGLYA 1570					
QY	220	TTAHADRVIEAEDDDSEFQQLHQRLARERENPPQPKLGATPISARFQPKLTAVAES 279					
Db	1571	SOAERRRMDLTARTGDSVVLYTYDR-----QDKPNPDHVAIA-----ATNHS 1613					
QY	280	VLEGDTTQSPKPSMLKSGAGVPTLAVTLDKGLQLAPDNPPALNTLLKOTLKDQTQ 339					
Db	1614	AFIGLDTQONREFQT-----VPIVAPGDRLTYDSNYG-----TCTDDCV 1655					
QY	340	HYLAHASSDQSHLLLDKNKHLFDIKSTATSYSVLNHSPGEIKGLAQAGTGSVSDG 399					
Db	1656	RLVTVHIDYTD-PDHTLID-----MHRG-PNDVR-----DN 1683					
QY	400	KSGKISLGSQTOSHNTMLSQPGEAHRSLTLGTIWOHPAGAARPQGESIRLH-----D 451					

Db 2629 TST-----GSGHSTDLTW--SNHVGGGNLLAIDAGDGLLKMGAICTAKHVIADIAGNL 2680  
QY 1345 MVATGHDVMPYMGKTKSAGNASDWLSAKIKIPDLRIGAAVSGTLOCTLONSLKFKLTE 1404  
Db 2681 RIESLODTHOYRSTDRSLGGS-----LTAGAGFGSGSANLNHQITRSDYASV 2726  
QY 1405 DELPGF-----IHGLTH-----GTLTP-AELLQKGBIHKOMKSGKSLTFSVDTSAN-- 1448  
Db 2727 TEOSGLFTGDGGVQVLTVGGTHLIGALITSNIAHNGL-NSLDTGTLILIONENHANYT 2785  
QY 1449 ---LDLRAGINLNE-----DQSKPNGVTKARVS----- 1472  
Db 2786 ATQVNLGGYSRNGGTGVTGDOQGAHAATAQVPGTTLPSHNGLSASPPSAMTARDSSHSTT 2845  
QY 1473 -AGLSANLAAGSRSTTSQFGSTTSASNNRP----- 1506  
Db 2846 YSGISOGALTIRDDTAQHAGLTHGTAETIATLNRDLTDTATSNALTPIPDEORINAGFD 2905  
QY 1507 -----TFLNGVGAGANL-----TAALGVYAHSTHEGKPVGIFPAFTSTNVSAALA 1551  
Db 2906 IVSGLQRETGTTFINRAAEADLKTROATAADHAAHPSN-----GFNDQOQRTLROQAIA 2960  
QY 1552 LDNR-----TSQISILELKRAEPTVNSDISELTSTLCKHF-----KDSATT 1592  
Db 2961 LTNEAHALKDAMPGPGTYRQITTAALAG---ASGNVSAASSDLAKHIVNYVOOOGATAI 3017  
QY 1593 KMLAALKELDDAKPAE-QLHIL-----QOHFSA-----KDV 1622  
Db 3018 GHVATGQLTEGSPHLHAALHALLACAGAAASQOHCSSGAGAAASSVLTGLFSDPRPDT 3077  
QY 1623 VGDERYEAVRNKLKLVIRQOAAADSHMELGSAHSTTYNNLSINNDGI-----VELLHK 1677  
Db 3078 AQDR--EAKRNLITSIVTGIAISIGHT-DPATATHAA-----IAAVDNWLAAYVQVQMLNE 3130  
QY 1678 HFDALPASSAKRLGEMNNDPALKDIKLOSTPFPSSASVSMELKDLGREOTEKAILDG 1737  
Db 3131 EFEEA---TEKEKRL-----EKEKRAKRE-IDARODKLTVDG 3166  
QY 1738 KV-GREEVGVLPQDRNNLR---VKSVSVSQSVSKSEGFNTPALLLGTSNSAAMSERNIG 1793  
Db 3167 LLKGLKESGI---NDTNGLEHLFLHPVDTVHELKGI---LTHPTLLQLGESAVOELLNKVS 3222  
QY 1794 TIN--FKYGQDQNTPR-----RFTL-EGGIAQANQVASALTDLKEGLE 1835  
Db 3223 RMSEALLVGGQHAQGLGDLGVSIVADVCFALAAAGTAKAAEITLGEAGISLSKDVLE 3279

RESULT 6  
E82589  
hemagglutinin-like secreted protein XP2196 [imported] - Xylella fastidiosa (strain 9a5c)  
C:Species: Xylella fastidiosa  
C>Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
C:Accession: E82589  
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen  
Nature 406, 151-157, 2000  
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
A:Reference number: A82515; MUID:20365717  
A:Note: for a complete list of authors see reference number A59328 below  
A:Accession: E82589  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-3442 <SIM>  
A:Cross-references: GB:AE004032; GB:AE003849; NTD:9107324; PIDN:AAF84995.1; GSPDB:GN001  
A:Experimental source: strain 9a5c  
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A  
Briñones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carret, H  
as-Neto, E.; Docena, C.; El-Dorri, H.; facincani, A.P.; Ferreira, A.J.S.  
submitted to GenBank, June 2000  
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm  
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig  
chado, M.A.; Madeira, H.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E  
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;  
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A

Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sa  
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Si  
M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.  
A:Reference number: A59328  
A:Contents: annotation  
C:Genetics:  
A:Gene: XF2196

Query Match 2.6% Score 247; DB 2; Length 3442;  
Best Local Similarity 18.5% Pred No. 0.0013;  
Matches 412; Conservative 292; Mismatches 815; Indels 708; Gaps 103;  
QY 3 LKSLGTEHKAHVHTAAHNPVGHGVALQOQSSSSPONAASAAAEAGKNGKMPRIHPST 62  
Db 1374 IDNLGTGLYGDHIAL-----HAQTLNTRDETSDGHTHAATAAQRDLDIGADTLRNTAN 1428  
QY 63 A---ADGISAHAQKKSFLRGCL---GPKKF--SRSAPOGQPGTTHSGKATLRDLLARD 114  
Db 1429 AMILSDGDAAT-----GATLDMALHATGTATLLDNRSATIDITGLNITTTLLNNI--RD 1481  
QY 115 DGETOHEAAAPDA---ABLTRSGGVKRRMDDMAGRPVKGSGEDKVPPTQOKRRHOLNNF 171  
Db 1482 N---VHIAHAPDVVTEARMEOPHW--RKNQPN-----GGSGNFRFTSNYDAHDI--- 1525  
QY 172 GOMRQTMLSKMAHPASANAGDLQHSPPHIPGSHHEIKEEVPVGSSTK-----A 219  
Db 1526 -----YYLNPADIKKDD-----PYITPDGOQIHRALVRLTPTOTSAYFYAGGLVA 1570  
QY 220 TTAHADRVETIAQEDDDSEFOQLHQORLARERENPPQPKLGATPISARFQPKLTAAVES 279  
Db 1571 SQAERRMDLTARTGDSVVLYYTD---ODKQNPDPHAAA-----ATNHS 1613  
QY 280 VLEGTDTTQSPKQSMILKSGAGVTPLAVTLDKGLQLAPDNPALNTLLKQTLGKDTQ 339  
Db 1614 AFIGLDTPOQNERFQT-----VPITVAPGDRLTYDSNYG-----TCTDDCV 1655  
QY 340 HYLAAHASSDGSCHLLDNKNGHLFDIKSTATSYSLHNSHPGEIKGLAQAGTGSVSDG 399  
Db 1656 RLVTWHDYTD-PDHTLID-----MHRG-PNDVR-----DN 1683  
QY 400 KSGKISLGSQTSHNKTMLSQGEAHRSLLTGIMOHAPAGAARPOGESIRLH-----D 451  
Db 1684 EKVR---DATRTTQODILNDAGAPALITG---GAMMIQDTLNRHVDLLAGD 1733  
QY 452 DKTHILHPELVGWSADKTHSOLSRQADGKLYALKDNRTLONLSKSSKSLVKIKSY 511  
Db 1734 QTIVGLPPHTKEKSDDEHKY-----KRVLLIDNRALQ-LS-----RTDTFHN 1776  
QY 512 SVDQRGQVA-----ILTDTPGRHKMSIMPDLSPESHISLSLHFADAHGGLHG 561  
Db 1777 STTYRGKVPWSNESRTPTTQIGRITSGCHQIAAQTLLNNVTDSTHAPETIQLHTYN 1836  
QY 562 KSELAQSAIVSGRLVVADSEGLKFAAIPKQGDGNELKMKAMPOHALDEHFGHDHQS 621  
Db 1837 PS---TOTLSVNGVITVTDTPSLHTVSLADNGPSAQOELTYIPDQSIITTP----- 1885  
QY 622 GFPHDDHGQNLALVK-----NN-----FROQHACPLGNDHQHPGWN 658  
Db 1886 -----NAPIRDPAAPAVTVTPTGPTLPLPNNSLFTLHPDAATLITDPRFTLGRP 1935  
QY 659 LTDALVIDNOLGHLHTNPPEHIELDGMHLGSLAQEGKLHYFDOLT--KGWTGAESDCKQ 716  
Db 1936 YTSADTQLHALGDHDT---LHKRLGDGYEQRILRE---OLAQTGRRRLDGYTDDDOQ 1988  
QY 717 LKGLGDAAYLLKDGVEVK-----RLNINOSTSSI-----KHGTENVFSLPHVRN 760  
Db 1989 YRALLDAGTVAKOHLRPGIALSADQLAQTLSDIVWLVOQDQVLPDGTTRTALVPRLYL 2048  
QY 761 KPEPGDALOGLNKDDKAQAMAVIGVKNYLALTEKGTIRSFQIKPGTQOLERPAOTLSREG 820  
Db 2049 RPTGTD-----LTPDGALLAAASTINAH-TFTNTGTIDARHL-----IDINANTMQOG 2097

Qy 821 -----ISGELKDIHVDHKNLYALTH-EGEVPHOP--REAWONGAESSSWHKL- 865  
Db 2098 GRLTADAIHITGDTFTLGQFKARGYLKVHAQGNFLASSTLRDATTQGTTRHHSVTELD 2157  
Qy 866 -----ALPOSESKLSLDMSHEHKPIATPDGSOHOLKAGGWAYAAAPERGPLAVGT- 917  
Db 2158 QOAGFTVTGPAYLGL-STDQAMTHNGVAINTGTD-----GYTSLNA--TGPLHLGTL 2208  
Qy 918 -----SCSQTFFNRLMOGVKGVIPSGSLVTKLSA----- 947  
Db 2209 TTHRSDDTTQWDPNRSRHSRIDTEYCTISITG-NGDIOLNSGQDINLRAATLHSTQGTITAL 2267  
Qy 948 QTGGMT-----GAERKVSSEKFERIRAYAFNPTMSTPRPKNAAY 988  
Db 2268 ATGNVTITHGDTIQTSDSHTKRSGLLNSRTTTHADQOQOQAIGSTLSADKVFVKGN 2327  
Qy 989 ATOHCWOGREGKLPLEYMOGALIKOLDANVR-----HNAQPQDLQSKLETLDLGERGAEL 1044  
Db 2328 IYVTSVHVSDAGTYMQAEHDLTLQAATHTTQSTYSHTTKQRL-----IRNGGASL 2379  
Qy 1045 -LNDMKRFRDELEQSATRSVTVLGOHQ-----VLKSGNEINSEFKPSP 1087  
Db 2380 TLGNQSORTSTTTATTTGSLIGATNGVNTLIAGHYQOIGSDVLSPHGDDIDHAKYD 2439  
Qy 1088 KALVQSNVNSRGDL-----SKSQQAQVHATPPSAESKLQSMGLH 1129  
Db 2440 -----IIQAHHTSOTTQATRSGLTVALSTPLIAGQAQTAQOMHAAARSGDPRLOALAGL 2496  
Qy 1130 FVSAGVMSHQGEIPLGRQRPNDKLTAKSLIIDLTVTIGELHAKAKLVSHDKPD 1189  
Db 2497 TTALGA-----KNTIDAVRQ-DPRALGGLNAS-----LTVG----- 2526  
Qy 1190 ADQIKQLROOFTLREKRYESPVKHYTDMGFTHNKALEANYDAVAFNAFKKEHGVN 1249  
Db 2527 -----RSTHSTTTTTSTAAGSNVTAGGNVH---ISATGCTASTLIQSDYRG-- 2574  
Qy 1250 LATRTVLESQGAELAKLNTLLSDGESMSFSRSYG-----GGVSTVFVPTLS 1300  
Db 2575 -DTMYLKADGDIAL-DAQNTV-----TNORNRGRSAGVAVNLGSGGTSAGLTAHAS 2628  
Qy 1301 KVPVPVIFGAGITLDRAVNLSPSTSGG--LNVSPGRD-----GGVSGNI 1344  
Db 2629 TST-----GSGHSTDLTW--SNSHVGGGNLLAIDAGDGLLMKGAIGTAKHVADIACNL 2680  
Qy 1345 MVATGHDVMPYMTGKTSAGNASDWSAKHKISPDRLRIGAASVTGLOGLTQNSLKFPLTE 1404  
Db 2681 RIESLQTHQYRSTDRSLGS-----LTAGAGFGSANLHQITRSDYASV 2726  
Qy 1405 DELPGF-----IHGLTH--GTLTP-ALLQKIEHQMKQSGKLTFSVDTSAN-- 1448  
Db 2727 TEQSGLFTGDGQYQLTVGGQTHLIGGAITSNSTAIHNL-NSLDTGLIIONIENHANYT 2785  
Qy 1449 ---LDRAGINLNE-----DGSKPNGVTARVS----- 1472  
Db 2786 ATQVNLGGVSRNGGTVGTDOQHAATATQVPGTTLPSHNLGSLASPPSAMTARDSSHST 2845  
Qy 1473 -AGLSANLAAGSRERSTTSGFGSTTSASNNRP----- 1506  
Db 2846 YSGISOGALTIRDDTAQHALTGTAAETIATLNRDLDTTATSNALTPIFDEQRINAGFD 2905  
Qy 1507 -----TFLNGVCAGANL---TAALGVAH-----SSTHEG 1532  
Db 2906 IVSGLORETGTFNNRAAEADLTKRQATAADHAAHNSGENDQORHARQATALNEA 2965  
Qy 1533 KPV--GIFPAFTSTNVSAALALNRTSQSISLELKRAEPVTSNDISELSTLGRHF----- 1586  
Db 2966 HAKDANGPGGTQYRIITALLAAG--ASGNVS-----AASNVSAASSDLAKHMIVNY 3015  
Qy 1587 ---KDSATTMMLAKKELDDAKPAE-QLHIL-----QOHESA----- 1619  
Db 3016 VOQOGATAIGHWVATGTLTGESPLHAALHALLACAGAAASQOHCSSGAQGAASVLTGL 3075  
Qy 1620 -----KDVVGDERYEAVRNLLKVLIRQQAADSHSMELGSAHSHTYNNLSRINDGI-- 1671

Db 3076 FSDPRPEDITQDR--EAKRNLITSIVTGIASGTGNT-DAATATHAA-----IAAVDNWNLAA 3128  
Qy 1672 ---VELLHKHFDALPASSAKRIGCEMMNDPALDKIILQLOSTPFSSASVSMELKDLGRE 1728  
Db 3129 KQVYMLNNEFEAA-----TEKEGRL-----EEEKVRAKWE-IDA 3164  
Qy 1729 QTEKAILDGVK-GREEVGVLFQDRNNLR--VKSVSVSQSYSKSRGFTNPALLLTSNSA 1784  
Db 3165 RQKLIYVDGLLGLKLGESGI--NDINGLEHLELHPVDVTHELGKI--LTHPTLLQLGESA 3220  
Qy 1785 AMSMERNIGTIN--FKYQDQNTPRRTLEGG--IAQANPOVASALT-----D 1828  
Db 3221 VQELLNKRVRMSEALLVGGDQHA--QQFGEDLGSVIADGVVALAAAGTFFKAAEILGEAGIN 3279  
Qy 1829 LKKEGLE 1835  
Db 3280 LSKDVLE 3286

RESULT 7  
H81193  
hemagglutinin/hemolysin-related protein NMB0493 [imported] - Neisseria meningitidis  
C:Species: Neisseria meningitidis  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 31-Dec-2000  
C:Accession: H81193  
R;Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, R.; Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.; Ric, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masigiani, V.; Pizza, M.  
Science 287, 1809-1815, 2000  
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.  
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.  
A:Reference number: A81000; MUID:20175755  
A:Accession: H81193  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-2703 <NET>  
A:Cross-references: GB:AE002405; GB:AE002098; NID:g7225708; PIDN:AAF40927.1; PID:g7272  
A:Experimental source: serogroup B, strain MMD58  
C:Genetics:  
A:Gene: NMB0493

Query Match 2.6%; Score 245; DB 2; Length 2703;  
Best Local Similarity 19.0%; Pred No. 0.0011;  
Matches 414; Conservative 282; Mismatches 766; Indels 712; Gaps 105;

Qy 10 HKAHVHTAAHNPVGHVALQOQSSSSSPONAAASLAAGKNRCKMPRIHQPSTAAD---G 66  
Db 573 HDLAVNTQTAKNSGH-LLTQTKIDNRELHNAGEIAA---NNLTIHSGRLSNDKKN 626  
Qy 67 ISAAHQOKKSFSLR-----GCLGTTKFSRSPQOGPTTHSKGATLRLDLDGGE 117  
Db 627 TRAAHLQDLTAGLNAGNILDGVTTKNNLNKNTQKVSVARLNTGQTLNTRGREAE 686  
Qy 118 TOHEAAAAPDAALTRSGG-----VKRRNMDMAGRPM-----YKGG-----SGE 156  
Db 687 TVN-----TQSQLTWQSGHITATEQLTINSRVNDQNGKLLSANQAOLAVSDGLYNOHGE 742  
Qy 157 DKVPTQOKKHQNLNFGQMOTMLSKMAHPASANAGRLQHPHPPHSGSHHEIKEEYPVST 216  
Db 743 IATNRQLSTHDKN-----QNTLALNNADGTIOSAGN-----VSLQAQSLA 782  
Qy 217 SKATTAHADRVETAOEDDDSEFQQLHQRLARENPQPPLKGVATPISARFQPKLTAV 276  
Db 783 NNGTLTAGNKLDTALTDD-----FVVERD-----LTAGQLNLS-- 816  
Qy 277 AESVLEGTDTTOSPLKPSQMLKSGAGVTPLATVTLKGLQLQAPDNPALNTLLKQTLGK 336  
Db 817 IKGRLLKNTHTLQ-----AGHT---LKLNAGNI---DN-----QVTGK 847  
Qy 337 DTQHYLAHASSDGSQHLHLLDNKGLHFLDKSTATSYLSVHNSHPGEIKGLAQAGTSVS 396

Db 848 -----IIGGEQDITSEQVHNRG-----LINS-----DGLTHIGAGQTL 882  
 QY 397 VDKSGKISLGSQTSHNKTMLSPGEGAHRSLLTGIWHPAGAAPOGE--SIRLHDDKI 454  
 Db 883 TNGTGKI-YGNHIALDAQILLNREETTEGTGKAG-----ATAARKRLDIGAKEITHNQBG 936  
 QY 455 HILHPE--LGWQWQADKDTSHO-----LSRADGKLYALKDNRTLONLSDNKS 500  
 Db 937 ALLSSEGI FAVGNRLDEQHAAGMADTFVNGSAGLEVQCD-----ALMSVRNMONINNHFK 992  
 QY 501 SE-----KLVDKIKSYV-----DQGOVALITDTPGRHKMSIMPSLDASPESHISLS-L 549  
 Db 993 TETYLAKAEKOVRYTYLQNTYTYAGKGLDFNSQGO-KQDTATPHLKNRSRTEANQW 1051  
 QY 550 HFADAH-----QGLLH-----GKSELEAQSVASHGLRVAD-----581  
 Db 1052 HVRDYHLETYKERIENRPAHITVGGDLTASGQWLNKDSRIVVGGRIITDOLNKEITN 1111  
 QY 582 --SEGKLFSAAI PKQD-----GNELKMKAMPOHALDEHFGHDIHISGTFHDDHCQ 630  
 Db 1112 QSTTGKRTDAVGTQWDSVTKKGYSGKRQRRTERNHP-----YHDTQL--FTHDEFTP 1165  
 QY 631 LNALVN-----NFRQOHC-----PLGNHQFHPG--WNLTDALVIDNQLGLHRTN 675  
 Db 1166 VSVIQNAASPOPAASAIAIKLDGVSTAAVNGQRIHTGNVSVLNNATVTLPNSSLYTH 1225  
 QY 676 PEPHEILDMLHGLSLALQEGKLHYFDQLTGWTGAESDCKOLKGLDGAAYLLKDGVEKR 735  
 Db 1226 PD-----LVETDPOFADY--RRWIGSDYMLQQLQDTHLHKLRLGDGYYEQ 1273  
 QY 736 LNIHQSTSSIKHGTENVFSLPHVRNKPPEPDALQGLNKKDKQAQAMAVIGNVNYLALTEKG 795  
 Db 1274 KLVN-----EQIHQLTGYRR-----LDGYSDEE-----QFKALMDNG 1306  
 QY 796 --DIRSFOIKPG---TOQ-----LERPAOTLSREGISGELKDIIHVDHKNQIYAL 839  
 Db 1307 LPAAKTFLGTPGIALSAEQVARTSDIVVMENQTVLS-----DGSTQTVLV--PKVYAL 1359  
 QY 840 THEGEV-----FHQPREAWONGAESSS-----WKLKALPQSES-----872  
 Db 1360 ARKGDLNTSGGLISAEQVLLKLQNLNTSGTLAGRAVLIOARNINSNGNIQADQIGLK 1419  
 QY 873 KLSLSDMSHEKPIATFEDGSQHO-----LKAGWHAYAAAPER-----G 911  
 Db 1420 AEKSNL-----DGOVQVQAGRLTLTAQAINLNGTITQTSNGERNGNTAIDRMAG 1468  
 QY 912 PLAVGTSQTVFNRLMOGVKGVIPGSLGTVKLSAQGTGMTGAEGKRYSSKFESE--RI 968  
 Db 1469 INVVG-SHTEQVDNRTSDGI-----LSLHASNDINLNAATVSNQVKDGTQI 1514  
 QY 969 RAYAFNPTMSTPRPIKNAAYATOHQWOGREGKPLYEMOGALIKOLDAHNVHRNAPQDPL 1028  
 Db 1515 TA-GNNLNLGTTIRTEHREAYGT-----LDDENHHRVROSTEV 1550  
 QY 1029 OSKLETLDLGHEGALL--NDMKRERDELEQ-----SATRSVT-----VL 1066  
 Db 1551 GSSRT-----QNGALRLRAGNDLIKROGELEABEGKTVLAAGRDVITISEGRQITELDTSVS 1606  
 QY 1067 GOHQGVLSNGEINSEFKPS-----PGKALV---QSFNVN-----RSQ 1101  
 Db 1607 CKSKGILSST-KTHDRYRFSDHEAVGNSITGGKMIYVAGQDINVRGSLNLSKGIYVLKAG 1665  
 QY 1102 QDLKSLSQOAVHATPPSAESKLSQMLGHFVSAGVDMSHOKGEIPLGRQR--DPNOKTALT 1159  
 Db 1666 HDIDISTAHNRYTGNEXHESKSGVMG---TGGLGFT-----IGNRKTTDDTDRNTIV 1715  
 QY 1160 KSRIL-----DVTYIGELHELADKALVSDHKPDADQIKOLRQOFDTLREKRYESNPVK 1214  
 Db 1716 HTGSIIGSLNGDVTYVAGNRVROTGSTVSSPEGRNTVTAKSIDVEF---ANNRYATDYAH 1772  
 QY 1215 HYTMGFT---HNKALEANYDAVKAFINAFKKEHGVNLTTTRTVLESQGSASLAKLKNKT 1271  
 Db 1773 TQEQKGLTVALNVPVQAAQNFIAQAQNVGKSKNRVNAANAANAQW-SYQATQOMQOF 1831

QY 1272 LLSLDGSEMSFSRSGYGGVSTVFVPTLSK-----KVPVPVIPGAGIITLDRAINLSF 1323  
 Db 1832 APSSAGOGNNQSPSISVITYGEOKSRNEQRHRYTEAASQIIGRGOT-----TL 1884  
 QY 1324 SRTSGG-----LNVSPGRDGGVSGNIMVATGHVDM--PYMTGKKTSGAGNASDW--LSAKHKI 1376  
 Db 1885 AATSGEQSNINITGSDVIGHAGTALIADNHIRLQSAKQDGSEQSKNKSNGNAGVAVKI 1944  
 QY 1377 SPDLRIGAAVSGTL-OGLTLQNSLKFKLTDELPCFIHGLTH-----GTLT 1420  
 Db 1945 GNGIRGFIYAGNIGKKEGG-----STTHRHVGVSTGTGKTIRSGGDTTLK 1993  
 QY 1421 PAELLOKQIE-----HOMKQ-----GSKLT-----1440  
 Db 1994 GVOLICKGIQADTRNLHIESVQDTEYQSKQONGNVQTVGYGFSASCSYQSKVKADHA 2053  
 QY 1441 -----FSDYTSANLDRAGINLNEDEGSKPNGVTVARVSAGLSASANLAAGSR 1486  
 Db 2054 SVTGSGGIYAGEDCYQIKVRDNTDLKGGIITSSQSAEDKGNLFTATLTAS-DIQNHRSR 2112  
 QY 1487 ERSTTSGQFGS-----TTSASNRPTEFLNCVGAG-----ANLTAALGV-----AH 1526  
 Db 2113 YEGRSFGIGSPDLNGWGDGTVDKQGRPTDRISPAAGYSGDGSKNSTTTRSGVNTNINH 2172  
 QY 1527 SSTHEKPGVIFPAFTSTNVSAALALDNRTSQSISLELKRA--EPVTSNDIS---ELTST 1581  
 Db 2173 ITDEAQLARTGRTAKETEARIYTGIDTETADQHSGLHKNKSPDKDAVAKELNLOREVTK 2232  
 QY 1582 LCKHKFDSATTKMLAALKELDDAKPAEOLHILOOHFSAKDVGVGDYERVAVRNKKLVIRQ 1641  
 Db 2233 FGNN-----AAQAVAAV-----ADKLGNTQSY-----ERYOEARTL--LEAEL 2268  
 QY 1642 QAADSHSM-----LGSASHSTTYNNLSRINN 1668  
 Db 2269 QNTDSEAEKAAFRASLGQVNAVLAENQSYDYDQWKEGIGRSILHGAAGLTTGSLGILA 2328  
 QY 1669 DGIVELLHKHFDALPASSAKRLGEMNNDPALKDIIKQLQSTPFFSSASYSMELKDLGRE 1728  
 Db 2329 GGGTSLAAPYLDKA-----AENLG-----PAGKAAVNALGGAAGVAT-----2366  
 QY 1729 QTEKAILDGKVGREBVGLFQDRNN--LRVKSVSVS-----QSVSKSEGTNTPALLL 1778  
 Db 2367 -----GGSGAVVGANV-DWNNRQLHPKEMALADKAYEALKREVEKREGRKI-----2412  
 QY 1779 GTSNSAAMSMEINI 1792  
 Db 2413 -SSQEAAMRIRROI 2425

RESULT 8

A56539

giantin - human

N:Alternate names: macrogolin

C:Species: Homo sapiens (man)

C>Date: 19-Oct-1995 #sequence\_revision 26-Jan-1996 #text\_change 10-Dec-1999

C:Accession: A56539, S37536

R:Seelig, H.P.; Schranz, P.; Schroeter, H.; Wiemann, C.; Griffiths, G.; Renz, M.

Mol. Cell. Biol. 14, 2564-2576, 1994

A:Title: Molecular genetic analyses of a 376-kilodalton Golgi complex membrane prote

A:Reference number: A56539; MUID:94187728

A:Molecule type: mRNA

A:Residues: 1-3259 <SEE>

A:Cross-references: EMBL:X75304; NID:9405714; PIDN:CAA53052.1; PID:g405715

C:Genetics:

A:Gene: GDB:GOLGB1; GCP; GCP371

A:Cross-references: GDB:454958

A:Map position: 3q13.31-3q13.31

C:Superfamily: giantin

C:Keywords: coiled coil; Golgi apparatus; transmembrane protein

F:3238-3254/Domain: transmembrane #status predicted <TMN>

Query Match 2.6%; Score 244.5; DB 1; Length 3259;  
Best Local Similarity 18.6%; Pred. No. 0.0016;  
Matches 389; Conservative 311; Mismatches 816; Indels 577; Gaps 92;

QY 23 GHGVALQOQSSSSPONAASLAAGKRNKMPRIHQ--PSTAADGISAHAQOKKSFSLR 80  
DB 603 GEGIA-----PIKMKVLEDTGDFPLMPNNESSLPVAVEKEQASTHQRTSEEI- 652

QY 81 GCLGKTKFSRSPQGPQGTTHSKGATRLDILLARDGETQHEAAPDAAR----- 129  
DB 653 -----SLNDAGVELKS--TKQDGD-KLSAVPDIGQCHODELERLKS 691

QY 130 -----LTRGGVKKRRNMDDMAGRPWKVGGEGEDKVPQOKRHQLNN-----FCQMR 175  
DB 692 QILELELNPHKAQEIYEKNLDERA-----KEISNLNQLIEFEKKNADNNSAFTALSEER 746

QY 176 QTMLSKM-----AHPASANAGDRLOHSPPHIPGSHHEIKEEPVGTSK 218  
DB 747 DQLLSQVKELSVYTELRAQVKQLEMLNLAERQRLDYES---QTAHDLNLLTEQIHSLSI 803

QY 219 ATTAAHADRVEIAQED-DDSEFOQLHOORLARENPPOPKLGV-----ATPLSARFQ 271  
DB 804 EAKSKDVKEIVLQNELDDVOLQFSEOSTLIRLSQSQLOQNKESVLEGAERVHRHISKVEE 863

QY 272 KLTAVAESVLEGDTTQSPKLP-----OSMLKSGCAGVPLAVTLTDKGLKLAPON-- 322  
DB 864 LSQALSQKELEITKMMOLLEKKRDVETLOQTTEKQDOQVTEISFMTKMWQLNEEKS 923

QY 323 -----PPALNTLLKQTLGKTQHYLAHASSDGSQHLNLLDKHGLFDIKSTATSYV 374  
DB 924 LGVEIKTLKEQLNLSRAEAKKEQVEEDNEVSSGLKQ-----YDEMSAPAGLSK 974

QY 375 LHNHPGEI-----KGLAQAGTGSVSDGKSGKIS---LGSGTQSHNKTMLSQP-- 421  
DB 975 EELQHEFDLLKKEQBRKRLQALINRKELLQVRLEELANLKDESKEIPLSETER 1034

QY 422 GEAHRSLLTGIWHPAGAPQGESIRLHDDKIHLHPCLGWQSAQDTHSQLSRQADG 481  
DB 1035 GEVEEDKENKEYSEKCVTSKQEIYLV---KQTISEKEVEL-QHIRKDLBEKLAARQF 1090

QY 482 KLVALKNRTLN-----LSDNKSSEKLVDKIKSVQ--QRQVAILTDTPGR 528  
DB 1091 QALVKQMNQTLQDKNTQIDLLQAEISENQ-----TIQKLTNTDASDGSVALVKET--- 1144

QY 529 HKMSIMPLDASPESHISLSLRFADAHQGLLHGKSELEAQAQVAISHGRLVWADSEKLF 588  
DB 1145 --VVISPPCTGSSE-----HWKPELEEKILALEKEK---EOLQKKLQE 1182

QY 589 A-----AIPKQGD-----GNELKMKAMPQHALDEHFGHDHQSISGFFHDDHQLNALYKN 637  
DB 1183 ALTSRKAILKKAQEKERHLEELAQKQDDYNRLQFQFDEQSKENENIGDQLRQLOQVRE 1242

QY 638 NF-----ROHACPLNGDHQFPGWNLTALVTDNOLGLHHTPEPEIILDMGHGLSL 690  
DB 1243 SIDGKLSTQDQESC-----SSTPG--LEPFLKATE--QHHTOP-----VLE----- 1281

QY 691 ALOEGKLYHFQDLTKGWTGAESDCKQLKGLDGAAY--LLKDGVEYKR-----UNINOSTS 744  
DB 1282 -----SNLCPDWPSSHEDASALOGGTSVAQIKALKEIAEKVELELKVSTSTSE 1331

QY 745 IKHGTENVFSLPHVRNKPDPGALQGLNKKDDKAQAMAVIGVKNYLALTEKGDIRSQIKP 804  
DB 1332 LTKKSEEVFOLEQELNK-----QGL-----EIES--LKT 1358

QY 805 GTQOLERPAOTLRREGISGELKDIHVDHKQNLIALTHGEVEFHQPRE---AMQNGAES 861  
DB 1359 VSHEAEVRAESLQKLESSQIAGLEHLRELQPKLDELQKLSKKEEDVSYLSGQLSSEK 1418

QY 862 WHKLA-----LPQSESKLSLSDMSHEHKPIATFEGSQHQLKAGGWHAAAPERGPLAYG 916  
DB 1419 EAAULTKIOTEITEQEDLIKALHTQLEMO--AKEHDERIKOLOVELCEMKQKPEE---IG 1472

QY 917 TSGSQTVEFNRMLQGVKGVIPGSGLTVKLKSAQTGTMGAERKVSFKFSERI---RAYAF 973  
DB 1473 EE-----SRAKQIQIRKL-----QAALISRKALKENKSLQELSLARGTIE 1514

QY 974 NPTWSTPRPIKAAAYATQHWGQGRE-----GLKPLYEMOGALIKOLDAHNVHNAPOPD 1027  
DB 1515 RLTKSL-----ADVESQVSAQNKEKDTVLGRLLALLOEERDKLITEMD---KSLLENOS 1564

QY 1028 LOSKLETLDLGHEGAELNDMKRFRDELEOSATRSVTVLQHQGVLSNGEINSEKPKSP 1087  
DB 1565 LSSCESKULALEG--LTEDKEKLVKEIESLKSSKIAESTEWQ---EKHKELQKEY--- 1615

QY 1088 GKALVQSENVNRSQDLSKSLQQAQVHATPPSAESKLQSLMGLHFVSGVDMSH-----QKG 1142  
DB 1616 -EILLOSQY-----ENVSNEAERIQHVVBAVROEK--QELYGKLRSTEANKKETEKOLEA 1667

QY 1143 EIPLRQRPDPDKTALTAKSRLLDIV-----TIGELHELADKAK-----LVSDHKPDADQ 1192  
DB 1668 EOEEMEEMKMKFAKSKOOKILELEENDRLRAEAVHPAGDTAKCEMETYLLSSNASMKEE 1727

QY 1193 IKOLROOFTLREK-----RYESNPVKHYTDMGFTTHNKALEANYDAYKAFINAF 1241  
DB 1728 LERVKMEYETLSKKFQSLMSEKDSLSEEVQDLKHQIE---DNVSKQANLEAT--- 1776

QY 1242 KKEHGVNLTTFTVLESQGSAAELAKKLNTLLSLDS---GESMSFSRSYGGGVSTVF-- 1295  
DB 1777 EKHDQTNVTERGTOSIPGETE-----EODLSMSTRPTCSSESVPSAKSANPAVKDFSS 1831

QY 1296 -----VPTLSKKVPVPVPIPGAGITLDRAYNLSFSRTSGG-----LNVSGRGGV 1340  
DB 1832 HDEINNYLOQIDOLKERI-----AGLEEKOKNEKFSOTLENEKNTLLSOISTKDQEL 1884

QY 1341 -----SGNIMVATGHDVMPYMTCKTSAGNASDWLSAKHKISPLDRICAAVSGTLOG 1392  
DB 1885 KMLQEVVTMNLNQIOEELSRTVKLKEAEEKD-----DLEE 1924

QY 1393 TLQNSLKFKLTEDELPGFTHTGLTTPAELLQKQIEHOMKOGSKLTFYSVD----- 1444  
DB 1925 RLQNQLA-----ELNGSI--GNVCQDVDAQINNELLESEMKNLKKCVSELEBEKQOLVK 1977

QY 1445 --TSANLDR-----AGINLNEGSKPNV-----TARVSAGLS 1476  
DB 1978 EKTKEVESEIKLEYLQIOGAQ--KEPGNKSHAKELQELLKEKQEVKQLOKQDCIRYQKIS 2036

QY 1477 ASANLAAGSRERSTTSQFGSTTSASNNRPTFLGVGAGANLTAALGVAHSTHEGKPVG 1536  
DB 2037 A-----LERTVKALEFVQTESOKDLEITKEN-----LAQAVEHRRKKAQELAS 2079

QY 1537 IFPAFTSTNVSAALAL--DN-----RTSQSISLELKRABEPTVTSNDIS--ELTSTLGHKF 1586  
DB 2080 FKVLDDTQSEAAARVLADNLKLLKELQSNKESVKSMQKQD---EDLERLEQAEKHL 2135

QY 1587 KDSATTMLAALKELDDAKPAEQHLIQQHFSKADVVGDERYEAVRNKLKLVIROOAAADS 1646  
DB 2136 KEKN-----MOEKDLALRREKVLH-----EETIGEIOVTNLKKDKVEQVQLOENLDS 2182

QY 1647 HSMELG--SASHSTYNNLSRINNDGIVELLKHFDALPASSAKRGLGEMNNDPALKDI 1704  
DB 2183 TVTQLAAFTKSMSSLODDRVR--IDEAKKWERKFSDAIOSKEEIRLKE--DNCSVLKQD 2239

QY 1705 IKOLQSTPFSASVSM--ELKDGL--REOTEKAILGKVGRE----- 1742  
DB 2240 LRQM-----SHMEELKINISRLHDKQIWESKAQTEVQLOKQVCDTLOGENKELLS 2291

QY 1743 -----EVGLFTQDRNNLRVKSVSQSQSKSGFNTPLLTLGTSNSAASM 1788  
DB 2292 QLEETRHLHYSSQNELAKLESELKSLDKQLDLDLSNLEKCKEQK-----GNLEGIROQ 2345

QY 1789 ERNIGTINFKYQ---DQNTPRRET--LEGGIAQANPOVASALTDLKKEGLEM 1836  
DB 2346 EADIONSKFSYEQLTDLQASRELTSLRHEEINNEKQIISLLSG--KEAIOV 2397





QY 1647 HSMELG--SASHSTYNNLSRINDGIVELLHKKHFDPAALPASSAKRLGEMMNDPALKDI 1704  
Db 2149 TVTOLAAFTKMSISQDDRRDY-IDEAKKWERKFSDAIOSKEEIRLKE--DNCISVLKDD 2205  
QY 1705 IKOLASTPSSASVSM-ELKDGIL-REOTEKAILDGVKRGRE-----1742  
Db 2206 LROM-----SIIMEFLKINISLEHDKQIWESKAOTEVQLOQKVCDTLOGENKELLS 2257  
QY 1743 -----EVGVLFQDRNNLRVKSVSQSVSKSEGNTPALLLGTGNSAASM 1788  
Db 2258 QLEETRHLYHSSONELAKSELKSLKQDLTDLNSLEKCKEQK-----GNLEGIIRQ 2311  
QY 1789 ERNIGTINFKYQ--DQNTPRFT--LEGGIAQANPOVASALTDLKKEGLEM 1836  
Db 2312 EADIONSKEFYQLETDLQASRELTSRLHEEINMKKEKIIISLSG-KEEAIQV 2363

RESULT 10  
T13564  
microtubule-associated protein homolog - fruit fly (Drosophila melanogaster)  
N:Alternate names: hypothetical protein EG:49E4.1  
C:Species: Drosophila melanogaster  
C>Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 17-Nov-2000  
C:Accession: T13564  
R:Spanos, L.; Papagiannakis, G.; Siden-Kiamos, I.; Louis, C.  
submitted to the EMBL Data Library, April 1999  
A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.  
A:Reference number: Z17689  
A:Accession: T13564  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-5327 <SPA>  
A:Cross-references: EMBL:AL031128; PIDN:CAA20006.1  
C:Genetics:  
A:Cross-references: FlyBase:FBgn0025392  
A:Introns: 24/2; 52/3; 104/3; 179/1; 232/1; 1669/3; 2566/1; 4798/3; 5272/1  
A>Note: EG:49E4.1  
C:Superfamily: Drosophila 576k microtubule-associated protein homolog

Query Match 2.5%; Score 240.5; DB 2; Length 5327;  
Best Local Similarity 18.4%; Pred. No. 0.0056;  
Matches 370; Conservative 271; Mismatches 803; Indels 567; Gaps 81;

QY 18 AHPVGHGVALQOGSSSSPQNAASLAAGKNGKMPRIHOPSTAADGI-----SAAH 71  
Db 3517 ASRPTSVAESVKDEAKSKESRRRESVAESKSLASK--EASRPASVAESVKDEAKSKEE 3574  
QY 72 QOKKFSRLGCLTKKFSRPAQOG---PGTHSKGATLRDLLARDGGETOHEAAPDAA 128  
Db 3575 SRRESVAESKSLASKESRPAASVAESVKDEAKSKESRRRESVAESKSLPSKEASRPPTS 3634  
QY 129 -----RLTSGGVKRRNMDM--AGRPVKGSGGDKVPTQOKRHQLNFGQM 174  
Db 3635 AESVKDEADKSKESRRRESVAESKSLASKESRRRESVAESVKDEAKSKESRRRESVTEK 3694  
QY 175 RQMLSKMAHPASA--NAGDRLOHSPHPHIGSHHIEKEEYPVGSSTSKATTAHADRVETIAOE 232  
Db 3695 SPLPSKEASRPTSVAESVKDEAK-----SKEESRRRESVAESKSLASKESRRSPASVAE 3747  
QY 233 DDSEFQOLHQRLARENPPOPKL-----GVATPISARQPKLTAVAESVLEGTDTTO 288  
Db 3748 SIKDEAGTKQE---SRRESPESGKAESIKGQDSSLASKETSRRPDSVRESVKDETE--- 3801  
QY 289 SPLKPO-SMLKSGAGVTPLATVLDKGLQALAPNPPALNTLLKOTLKGDKTOHYLAHHS 347  
Db 3802 ---KPEGAIDKQVARSPEVAVS-----AKDEKSLPHSRPESVADKSPD-----AS 3846  
QY 348 SDGSQHLLLDKNGHLDFIKSTATS-----YSVLHNSHP-----GEIKKLAQAGTGSVSV 397  
Db 3847 KEAKRSL-----SVAETASSPIERGPRSIADLSPLNLTGEAKGLP---TLSSPI 3894

QY 398 DGSGKISLGSQTSQSHNKTMLSQPCEAHRSLLTGTWQHPAGAAARPOGESIRLHDDKIHL 457  
Db 3995 DVAEGDFLEVKAESSPRPAVLSKPAEFQPDGTGHTASTPVDSEAPVLEEIEVEQ----- 3949  
QY 458 HPELGWQVSADKQTHSRLSQADKLYALKDNR--TLQNLSDNKSSEKILVDIKYSVDO 515  
Db 3950 HTTSGV-----GATGATAETDLDTETKSETVTVKQSETTLFETLSKVES----- 3995  
QY 516 RGQVAILTDTGPRKMSIMPSLDASPESHISLSLHFA-----DAHQGLLHGKSE--LEAQ 568  
Db 3996 -----KVELESVQVKEKQVTSVKAQETVTDLSLEOLFCKSSEQLTEIK 4041  
QY 569 SVAISHGLRVADSEGLFSAAI-----PKQDGNELKMKAMPQHALDEHFGHDHQLSGFF 624  
Db 4042 SVLDTN-----ISNVTNLFSTAVETIEKKVQDVTEKVIKATEH-VSEHVTGTTGESSTET 4095  
QY 625 HDDHQQLNALVKNFRQOHCAPLGN-----DHQFHPGWNLT 661  
Db 4096 SOEKSSLDLGTSELRETHITTVGSPFTVTICERDEPVLHDIKEDEEHFSPSPDVK 4155  
QY 662 ALVIDNQLGLHHTNPPEHIELDMGLSLALQEGK-----LHYFDOLTGTGWTGAESDCKQ 716  
Db 4156 AALIPPQ-PMRPLSPREEVAKIVADVAKVLKSKDKDITDIIPDFE-----RQ 4202  
QY 717 LKGLDGAAYLLKDGVRNLNQSTSSIKHGTENVFSLPHVRNKPEPGDALQGLNKDDK 776  
Db 4203 LEEKLKSTADTEESDKSTRDEKSLSEISVKVEIESKSPDQKSGPISIEKKDIOSEK 4262  
QY 777 AQAMAVIGVKNYALTEKGDIRSFQIKPGTOQLERPAOTLSREGISGELKDIHVDHKONL 836  
Db 4263 A-----QLRQILASSRP-----ESVASQPSVSPSQS-- 4291  
QY 837 YALTHEGEVHOPREAMQNGAESSWHKLLALPQS--ESKLSLDMSEHKPIATFEDGSOH 895  
Db 4292 -AASHE-----HKEVEL-----SESHKAEKSSRPESVASQVSEKDKMKTSRPASST-----SQF 4338  
QY 896 QLKAGGWHAAYAPERGLAVGTSGSOTVNRMLMOGVKGVIPGSLGTVKLKSAOTGGMTGA 955  
Db 4339 STKGEDEETES-----LLHSLTTTETVETKQME-----EKSFSFVSSTSVT 4380  
QY 956 EGRKVSSEKFSERIRAYAFNPTMTSPRIKNAAYATQHGMOGREGKLPKLYEQMALIKOLD 1015  
Db 4381 KSTVLSSQSTVOLREESTSESLSSSLKVEDSS-----RRESLSLLAEKGI----- 4427  
QY 1016 AHNVRHNAPOPDLQSKLETLDLGHECAELLNDMKRFRDELEQSATRSVTVLGQHQGVLS 1075  
Db 4428 ATNTSLKEDTSASOLE-----ELL-----VQSEECSSSESIVSEIQTISIAKS 4471  
QY 1076 NGEI-----NSEKPSPGKALVQ-----SFNVNRSGQD-- 1103  
Db 4472 NKEIKDARETKVTSQTTTTSSATKDDSLKETVAEFLATEKIVSAKEAFSTEATKSADD 4531  
QY 1104 LSKSLQAAVHAATPPSA-----ESKLQMLGHFVSAGVDMHSHQKEIPLGRORDPNDK 1155  
Db 4532 LAKTTASAVSSTASQALFVGTDESRRESLLSQAESRLTHSDPEDEPAD-----DVER 4588  
QY 1156 TALTKSLRLDVTIGELHELADKAKLVSDHKPDQAIKOLRQOQDTLREKRESNPVKH 1215  
Db 4589 SVSKESR-----SKSIATI-----MMTSIYKPSDEM-----EPI-- 4617  
QY 1216 YTDMGPTHNKALPANYDAVKAFINAPKKEHGVNLTTRTVLES--QGSAAELAKLKN--- 1270  
Db 4618 -----SKLVEEHEHVE-----ELAQEVTSSTKTT-TLIQSEQSTTTTSSTSKTGA 4664  
QY 1271 -----TLLSLDSESMSFSRSYGGGVSTVPVPTLSKKVPVPIPGAGITLDRAYNLSPSR 1325  
Db 4665 RVESITLTQMD-----QOTSQSGDPADRKTPPT-----APVSPGV----- 4700  
QY 1326 TSGGLNVFGRDGVGSGNINMVAATCHDYMPYMTGKTSAGNASDWLSAKHKISPDLRIGAA 1385  
Db 4701 -----KAMSSTG-----SAGSV-----IGAG 4716  
QY 1386 VSGTLOGTIONSFKFKLTDELPGFIHGLTHGTLTPAELLQKGIHQHMQGSKLTFSDVT 1445



Db 4717 AGAAGGCKESSAASIVS-----SSGPMSP-----KDISGKSPGALTSQSQI 4761  
 QY 1446 SANLDLRAGINLNEDEGKPNQVTAARVSAGLSASANLAAGSRERSTTSQFGSTTSASNNR 1505  
 Db 4762 PTLGRESHDTDPSSPKPTSPFRVRSKDELKSLQEMQHHSQELAGAAAAGACEGDI 4821  
 QY 1506 PTF--LNGVGAGANLTAALGVASHSTHEGKPVGIFPAFTSTNVSAALDNR---TSOSI 1560  
 Db 4822 PELHELRLG-----ECTTALS---GSTDK-----IITTTITVTVKVISADGKEIVTEOKT 4868  
 QY 1561 SLEKRAEPTVSNDSIELTSLGKHFKDSATPKMLAALKEL-----DD-----AKP 1606  
 Db 4869 VTTTDSSEPDKVVTTRTTSERDQLLPKEVALLRLGYRASTPQSEDEDELLGLSP 4928  
 QY 1607 AEQLHILQQHESA-----KDVVGDEREYAVRNKLKLVIRQQAADSHSMELGSAHSTT 1659  
 Db 4929 RSATVELQSSSGVSKRSOLDADGDE-----SODDIPPOYGSEHSTA 4972  
 QY 1660 YNNLSRINNDGIVELLHKHFDAAALPASSAKRLGEMMNNDPALKDI IKOLOSTPF----- 1713  
 Db 4973 RSILLPRTADPMA-----TSFYGALPOSF-----DVMKPSSTEPIDGAP 5013  
 QY 1714 ---SSASVSMELDGLREQT-----EKALDGKVGREVGVLFDQRNNLRVKSVSQ 1763  
 Db 5014 SGOSSQSESVES---SQTWAGHKFLDQADKDFQRALEE---HVQARGAEVMSVSTAKY 5066  
 QY 1764 SVSKSEGFNTPALLGTSN-----SAAMSERNIGTINFKYGQDONTPRFTLEGG 1814  
 Db 5067 SYSPSKAEEMEQIVSTAERQRPPLSDVQARVAESGFATVGSVASOQOQOEK-----GGE 5122  
 QY 1815 IQAANP-----OVASALTDLKKEGLE 1835  
 Db 5123 VEQAVPTTAVTASTATATASSTGALPKORLE 5153

RESULT 11  
 T31102  
 filamentous hemagglutinin 1 - Haemophilus ducreyi  
 C:Species: Haemophilus ducreyi  
 C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999  
 C:Accession: T31102  
 R:Ward, C.K.; Lumley, S.R.; Latimer, J.L.; Cope, L.D.; Hansen, E.J.  
 J. Bacteriol. 180, 6013-6022, 1998  
 A:Title: Haemophilus ducreyi secretes a filamentous hemagglutinin-like protein.  
 A:Reference number: Z20984; MUID:99030326  
 A:Accession: T31102  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-4152 <WAR>  
 C:Cross-references: EMBL:AF057695; NID:g3929017; PID:g3929018; PIDN:AAC79757.1  
 C:Genetics:  
 A:Gene: lspa1

Query Match 2.5%; Score 237.5; DB 2; Length 4152;  
 Best Local Similarity 18.9%; Pred. No. 0.0052;  
 Matches 345; Conservative 244; Mismatches 636; Indels 599; Gaps 88;

QY 350 GSQHLLDNKCHLFDIKSTATSYSLHNSHP-----GEIKGLAQAGTGSVS 397  
 Db 835 GGOGLNLTAKGNITN-DSNSTAIVLHNSNDINLANNKVVYIGIYSQ-----AGNISV 888  
 QY 398 DGK-----SGKISLCSGTQSHNKTMLSDGGEAHRSLLTGIMWHPAGAARPOGESIRL 449  
 Db 889 EAKLLHNDVKLSGNIT---TTTKSGNATVKT-----NSIGGLHDANSIRV 931  
 QY 450 HDDKHILHPELGVQWQADKTHSRSQADGKLYALKDNRFTLQNLSDNKSSEKLVDKIK 509  
 Db 932 GELTLN-----GKFADLD--NQLKVALRGKIYA-GSNLTFF-----AKEGEKQKSTA 976  
 QY 510 SYSVDQRGQVAILTDTGRRHKMSIMPSLDASPESHI--SLSLHFADAHQGLLHGKSELEAQ 568

Db 977 QAKIINRGTINV-----KNKLEYSNVDV---ENNMRSMQVNL---YEKIFNGDNPI--- 1022  
 QY 569 SVAISHGRVLVAVSESGKLFSAAPKQGDGNELKMKAMPQHAIHDEHF-----CHDHQISGF 623  
 Db 1023 TLLKNGVTFKAFDPSNRRRAS---NDGEGTNKKTFDNVVAHLIEAFSGSYNGNDHRAS-- 1078  
 QY 624 FHDHGHQLN-----ALVKNFRQOHACPLGNDHGFHPCWN-LT-----DALVID--NQ 668  
 Db 1079 ---DDGHVRSPPYLLVLAQAVNTEGENYLLKTALOHIFGPNWDLTTTNNDTTINDKWNQ 1135  
 QY 669 LGL-----HHTNPEPH---BILDMGHLSLALQEG-----KL 697  
 Db 1136 LAKLWEKFKNCCENNHSINLIYPADRGVEKAKIFAGVLRNGTGVEDKVVYQELNDRKAK 1195  
 QY 698 HYFDQLTGKTGWAESDCOKLKG-LDGAAYLLKDG-----VKRLNINQS 741  
 Db 1196 EYEDFAKFKQGRFK---SRFQNGEFWDGWAKEGNSYSGKETEEKYNGIKKEHTVNIG 1253  
 QY 742 TSSIKHGT---ENVSLPHVRNKPFGDALQGLNK-----GIDKSIISSELLAQPIYVAKADVPDPRV 1307  
 Db 1254 KHEIKVPTVSFENLNINHOODKSD-----TEKGDIRSFOIKPGTQQLERPAQT 815  
 QY 774 ---DDKA-----QAMAVIGVKNYLAL-----RIGDNYFEH--QL 1364  
 Db 1308 AONDKAVDEGLYRTRLSVINONNYLGAKYFFNQLDTEDDKLKGIK--RIGDNYFEH--QL 1364  
 QY 816 LSREGISGELKDIHVDHKONLY--ALTHE-----GEVEHQPRAWONGA 857  
 Db 1365 ITR--LIEKAVDNHLTKHGLHDIALVKLIDSASTQAKDLNLKVGAL--TEKOKNLK 1420  
 QY 858 ESSWH-----KLALPOSESKLSLMSHEHKPIATFEDGSQHLKAGWHAYAAPE 909  
 Db 1421 EDIWMVVKTEVNAQEVLPVPOVYLAKOTIEBEVEKQRGVGT-----GOIRAGIIDVKVDDV 1474  
 QY 910 RGP-----LAVGTSGSQTVFN-----RLMOGVKCVIPGSGTLVKLSAQGTGTMCAEG 957  
 Db 1475 RNTGTIAGYAVGLEAKNKLKNTGDIILSQRLSKLVGKKGLESTGVT---YVDETG---ATKV 1529  
 QY 958 RVSRSKFSERIRAYAFNPTMTSPRIKNAAYATQHGMOGREGLKPLYEMOGALIKQLDAH 1017  
 Db 1530 RKARIKSEGI-----YLETDKD-KNVD-LTASELKNGTG-----QIKAK 1567  
 QY 1018 NVRHN-APQPDLOSLETLDLGEHGAELNDMKRFR-----DELEQSATRSVTV 1065  
 Db 1568 DLNLDNIYETSYKYKYEKL-TGKNGEIGDRVITQTSQAQSVGTDAFHLHLSLEGDVNQ 1626  
 QY 1066 LGQHOGVLKSNGEINSEFKPSPKAL-----VQSFNVNRSGQDLKSL-OQAV 1112  
 Db 1627 TGSNLKANRTTCVVKGDFNTRKAGKDLFHRQIDVTSGTVYSASAGSGQSAGISLTDQGV 1686  
 QY 1113 HA-TPPSAESKLQSMGLHFV---SAGVDMSHOKGEIPLGRORDPNDKTLTKSRILDT 1167  
 Db 1687 ETYTNKTATAGANADVTNFMKRTRETETSLTHRNSEF-----NALSGELYVMGK 1735  
 QY 1168 VTIGEL-----HELADKALV-----SDHKPDA 1190  
 Db 1736 ADIGGVDINRDVEVIKTPETIAAEQAEEAKKAEKENEASETAAKETEENENVAEK 1795  
 QY 1191 DQTK-----OLROQFDTLREKRVESNPVKHYTD--MGFT-----HNKA---LEA 1229  
 Db 1796 DKTQPKFKLTDEEIAAAAFETKGEDFFAAYKAREEDRKKCFTLTSAEQIESTKARDEKT 1855  
 QY 1230 NYDAVKAFINAFKEH---HGVNLTTRTVLESQGS-----A 1262  
 Db 1856 TYELKVGVCAGAEAAHSAADAISNKAQIITDQNGLKQDGTVALQEAASDVNLATGDIA 1915  
 QY 1263 ELAKKLNTLLSLD--SGESMSFSRSYGGG-----VSTVFVPTLSKSKVPVP 1306  
 Db 1916 GASAKLKFELSTIEKKSRGASDGRSILGRLNLAARGCDITLNNVETTENSLSLKARDN 1975  
 QY 1307 VIPCAGIT--LDKAYNLFSRTSG---GLNV-----SFGRDGGVSGNIMVATGHDVMPYM 1356  
 Db 1976 VNVNSGVTEQKDESNQSOLKVTAGASSGCGVMAGGCSAGVSAGVSGS-----YN 2024

```
QY 1357 TGKTSAGNASDWSAKHKISPDLRIGAAVSGTLOCTLQNSLKFKLTELDELPGFIHGLTH 1416
Db 2025 ESNTTESTHTNLLRGK-----SLRVEA-----GKDFNLSSNVVDVHLHLDDVKGDN 2072
QY 1417 GLTPAELLOK--GIEHOMKOGSKLTFSDTSANLDRAGINLNEGSK-----PNGVTAR 1470
Db 2073 VYSKODYSYRKERYNSVYAGVGTAGCARPNSVGLGSAENENSKIVKQOAGISAK 2132
QY 1471 VSAGLSASANLAAG-----SRERSTTSGOGFSTTSASNNRPTF 1508
Db 2133 RTITGETNNLNLTCGYTENKGNPDELNVKGDITTHELKDEHHKGGSGVGVSETGVTO 2192
QY 1509 LAGVGAGANLTAALGVAHSS---THEGKPVGIFPAFTSTNVSAALALDRNTSOSISLEL 1564
Db 2193 VVNGRVRQKHYEATQHSISGINTKGTGVNF-----KDRSQSTEV 2236
QY 1565 KRAEPTVS-----NDISELTSTLGHKFKDSATTMLAALKELDDAKPAEQHLHQHF 1617
Db 2237 HRDDTTAATNFNPFELGDAELAKK-GKEKWDNRSKTTT----- 2274
QY 1618 SAKDVVDERYEAVRNKLKLVIRQAA-----DSHSMELGSAHSSTYNNLSRINNDGI 1671
Db 2275 SODSAHDRSRSEVNGYSELPRKATADNAGVDSPLIKGEAQOQL-ALIKAGNDVI 2333
QY 1672 VELLHKHFDAAALPASSAKRLEMMNDPALDKIIOLOSTPFFSSASVSMELKDLREOTE 1731
Db 2334 -----PEVQSLTQKARQSLVDESPIAETPALTRPOVK 2366
QY 1732 KAILDGKVGREVGVLFDQRNRLRVKSVSYQ---SVSKSEGFNTPALLLGTSNSAAMS 1788
Db 2367 SNT-----AESIEVPQRTKVSQDGEYAEITPPT-NKAALSS 2404
QY 1789 ERNIGTFNKYQDQNTPRRFTLE 1812
Db 2405 QDVG-----DTPTPRALRLE 2420

RESULT 12
A35140
hemolysin A precursor - Proteus mirabilis
C:Species: Proteus mirabilis
C:Date: 03-Aug-1990 #sequence_revision 03-Aug-1990 #text_change 08-Oct-1999
R:Uphoff, T.S.: Welch, R.A.
J. Bacteriol. 172, 1206-1216, 1990
A:Title: Nucleotide sequencing of the Proteus mirabilis calcium-independent hemolysin ge
A:Reference number: A35140; MUID:90170827
A:Accession: A35140
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1577 <GPH>
A:Cross-references: GB:M30186; NID:g150888; PIDN:AAA25657.1; PID:g150890

Query Match 2.5%; Score 232; DB 2; Length 1577;
Best Local Similarity 18.9%; Pred. No. 0.002;
Matches 329; Conservative 238; Mismatches 614; Indels 564; Gaps 84;

QY 261 VATPISARFQPKLTAFAESVLEGDTTQ-----SPLKPOSMLKGGAGVTPLAVLTDKG 314
Db 197 IAPRIDSR--GKITAETSAFTGQNTFSOHFDILSQKPVSAALDSYFFG-----SMQSG 248
QY 315 KLOLAPDNPALNLLKQTLGKDTQHYLAHASSSDGSOHLLEDNPKGHLFDIKSTATFSYV 374
Db 249 RIRI-----INT-----AEGSVKLAGKFTADNLSVKADN-----IQDTSQV 286
QY 375 LHNHPGEIKGLAQAAGTGSVSDGSKGISLGSQTQSHNKMTLSQPGEAHRSLLTGIWQ 434
Db 287 RYDSYDKD-----GSENYQYRGGITVNNSSQTLTKTELKKG--NITLVASSH 334
QY 435 HPAGARQGESIRLHDDKHILHLPGL-----VWQ-----SADKDTHSQLSRQA 479
```

```
QY 1476 S-----ASANLAAGSRERSTTSQFGSTTSASNNRPTFLNGVGAGANLTAAL 1522
Db 1191 GKMTDASSVNGLSAQFAIGKODEKSVSREGGTINSGN-----L 1232
QY 1523 GVAHSSTH-ECKPYGIFPAFTSTNVSAALADNR-TSOSISLELKRAEPTVTSNDISELTS 1580
Db 1233 TINGSVHLQGAQVN-----SKDTOLTSQSDIEITSAQ---STDYKKNWG 1275
QY 1581 T-LGKHFKDSATYKMLAALKELDDAKPAEQHLHL-----QQHPSAKDVVVGDERYE 1629
Db 1276 TDJGFNGKKTNTPT-----KEVTEKPKATSTHNIGGKLLVNVEDQOKTSHONATLETGTL 1330
QY 1630 AVRNKLKLV1--RQOAAADSHSMELGASHSHTYNNLSRINDGIVLELLKHKHFDAALPASS 1687
Db 1331 TINSKDLTSLGANTADSVTVNGVGSNLASQKESDRHVTVG-VNVGYNHTND--PKSS 1387
QY 1688 -----AKRIGEMNNDPALKDII-----KQIQTSPFSSASVSMELKGLREOTEKAIL 1735
Db 1388 QVNKTAKAGGSL--EKTIKDTIDSGIKSSSTDAISDKVNSLSSTIADTKGISDET-KAKI 1444
QY 1736 D---GKVG-----REEVGVLFQDRNNLRVKSVSQSVSKSEGTNPALL 1777
Db 1445 DQGFQGVNGIKNIVTGAEGHTANADIKVTHVD-NDAVTKTTSNNLDSLNVNGSTKL 1503
QY 1778 LGTS-----NSAAMSERNIGTINFKYGODQNTPRFTLEGGIAQAQPOVASALT 1827
Db 1504 TGAIEVVSQSQVDLGGSSVKLE-NIEGHHYEAGADLD-----LKSSVV 1545
QY 1828 DLKKE 1832
Db 1546 DLAKQ 1550

RESULT 13
T09083
hemagglutinin/hemolysin-related protein NMB1214 [imported] - Neisseria meningitidis (gro
N:Alternate names: probable secreted protein
C:Species: Neisseria meningitidis
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 31-Dec-2000
C:Accession: T09083; B81109
R:Simpson, N.J.; Spratt, B.G.
submitted to the EMBL Data Library, October 1997
A:Reference number: Z16558
A:Accession: T09083
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2273 <SIM>
A:Cross-references: EMBL:AF030941; NID:g2623257; PID:g2623258
A:Experimental source: strain 44/76
R:Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
ri, H.; Qin, H.; Vamathevan, J.; Gil, J.; Scarlato, V.; Maignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: A81000; MUID:20175755
A:Accession: B81109
A:Molecule type: DNA
A:Residues: 1-2273 <TEU>
A:Cross-references: GB:AE002469; GB:AE002098; NID:g7226446; PIDN:AAF41596.1; PID:g722645
A:Experimental source: serogroup B, strain MMD58
C:Genetics:
A:Gene: pspA; NMB1214

Query Match 2.4%; Score 229.5; DB 2; Length 2273;
Best Local Similarity 18.9%; Pred. No. 0.0048;
Matches 419; Conservative 285; Mismatches 790; Indels 725; Gaps 107;

QY 13 AVHTAAHNPV-CHGVALQOGSSSSSPONAAASL--AAEKNRKNKMPRIHPSTAAAGTSA 69
Db 66 ALGVAAPAPASGII---ADKSAPKNOQAVILOTAG-----LPQVNTQTPSSQGVSV 116
```





QY 1356 MTGKTSAGNASDMLSAKHKISPDLRIGAAVSGTLOQTIONSFKPLTEDELPFGFIHGLT 1415  
Db 1806 QT--RSHADVSD--SLDLSFG-----EDQLNSTSERCNEEPATSSVHASS 1848  
QY 1416 HGTLTPAELLQK--GIEH-----OMKOGSKLTFVSDTSANLDRAG----- 1454  
Db 1849 PDSLISGOLPKKVESLESLEYFTPIPTRAQSKLESIGSGLDLSOSSKKTQSARRTMOI 1908  
QY 1455 INL-----NEDGSKPGV-----TARVSAGLSASANAAGSRER 1488  
Db 1909 INITMTKTKKEEPESANTSFYLSRAPSVQSLHLQNPRAGRPPAASAPALASLFSQES 1968  
QY 1489 STTSQFGSTTSASNNRPTFLNGVGAGANLTAALGVA--HSSTHEKPGVIFPAFTSTNV 1546  
Db 1969 LAKTEHFSSDLSL--NNLPQYQHPTRRSARLSQTTGGRSFYMTSCODEP---DPQDDWTRI 2024  
QY 1547 SAALALDNRTSQ---SISLELAKRAEPTVTSNDISELTSTLCK-----HFKDSA 1590  
Db 2025 -AELQARNKTCPPHLKTSYPLESRPSIFSTIDEVKGDPKETLRRATLLPCQIQDSM 2083  
QY 1591 TTKMLAALKELDDAKP-ABQL--HILQOHFSAKDVVVDERY-----EAVRNKLKLVROQ 1642  
Db 2084 TSTRQTL-----AVPGAHLKGHNISTROQMKRVSEESHYGPDTPEAKKTATCFPRPMT 2138  
QY 1643 AADSH-----SME-LGSASHS-----TNNLSRINNDGIVELLHKHFDAAAL 1683  
Db 2139 PKDKHDTRRVSTMESKSGSSHQOAPTRQATAPSFINTPKLGNLSLLKRLNK---KTT 2195  
QY 1684 PASSAKRLG-----EMMNDPALKDIIKO--LOSTPFSSASVSMELKQGLREQTEK 1732  
Db 2196 PKNSPRGRGANGSTSTSNKPSHLSLRKSPQRSVSTAKSPRASNKLFERKQO 2250

Search completed: June 5, 2001, 18:18:11  
Job time: 237 sec

QY 505 VDKIKSYSDQGOVAAILTDTGRHKMSTMPISLDSAPSE-----SHISLSLHF 551  
Db 873 LEKINQLEGELSAANACIKTERAEK-KLVSALHSEKLIKIAQYGESERLSHLETAL-- 929  
QY 552 ADAHOGL-----LHGKSELEAQSVAI-----SHGRLVVADSEKGLFSAATPKQGDGN 598  
Db 930 SNAKQDLCLAKELSEDEKYKKAEEFAMVYLKEONSERIASLESELK--NSLAVVYKERKCE 988  
QY 599 ELKMKAMPQHALDEHFGHCHQHSQFPHDGHQNL--ALVKNFRQOHACPLGNDHOPHPG 656  
Db 989 SEKLSGEVEHL-----KRQL-----DSSQRHKEALAKNIEIKO---LINAKE----- 1029  
QY 657 WNLTALVDINQLHHTNPEPHEIILDMHGLSLALQEGKLHYFDQTLTKGTWGAESDCKO 716  
Db 1030 -KATSDLAIKSEMG-----AQLQKAVDTHKSEFSALO-----NE 1062  
QY 717 LKKGDLGAAVLLKDGVEKRLN-----INQSTSIKHGHTENVFSLPHVRNKPPEGDA 767  
Db 1063 LSRSLDLA--LKEGEVERLNKEAALRQEELOQOQQTITKLTEETALAAKDKV---A 1116  
QY 768 LQGLNKDDKAQAMAVIGNKYLA-----LTEKG-----DIRS-----FOIK 803  
Db 1117 LQ--EKEIKQOVATKGAEMAKLKSIVSEKSKRIECLIODIONOKRDLSCIOEOHOSK 1174  
QY 804 PGTOQ-----LER-----PAQTLRSREGIS----- 822  
Db 1175 LGSQGLQALITADLEKKCKEOKELICEPAONKAAEAKTLASEKASVERQLEGIOALEI 1234  
QY 823 GELKDIIHVDHKNL---YALTHE-----GEVFH--QPREAWQNGAESSSHWKL--- 865  
Db 1235 GKERQKACDLQKQLELWAVQEEKETLOALKKELFKHVQELEQSOTSFDTSSGEBALLYL 1294  
QY 866 -----ALPOSESKLSLDMSEH-----KPIATPE-----DGSQHL 897  
Db 1295 SEAQEROALTEAQAEOYQKEIEMKNKEVNSLOAEIKILSSKVTTEEVSVDFEQRLL 1354  
QY 898 KAGGWHAYAAPERGLAVGTSGSQTVFNRKM-----QGVKGV-----IPGSG 940  
Db 1355 KETSKAKLEKMKQLHMELEAS---FKELLEKNCAIDCLTTEAQNLAKEADQQRMAVDS 1411  
QY 941 LTVKLSAQGTGWTGAERKVSCKSERIRAYAFNPTMTSPRIKNAAYATOHGMOGREGI 1000  
Db 1412 LOOKLSSKA-----ETNHTLOOEIOAWO-----KNCAREQQICSLQOINI 1451  
QY 1001 KPLYEMOGALIKOLDA--HNVHRNAPQPDLOQSKLETLDLGEHGAELNDMK---RFRDEL 1055  
Db 1452 K-----SNOSLLEEFASLKHYSQEIIAERDLMOE-----KHQEELLSHKLTERRFOAEL 1500  
QY 1056 EQSATRSVTVLQHQGVILKS-----NGEIN--SEFKPSPGKALVQSFVNVNRSQDGL---SKSL 1108  
Db 1501 EKAKEDMTEIV-----LLKEKLHNOELQLHFKQSENSYSLTQISHLQOVNSQLLGANQSL 1555  
QY 1109 QAVVIATPPSABSKLQSMLGHEVVSAGVDMSHOKGEPIGR-----ORDPNDKTALTCSR 1162  
Db 1556 SQISQGAKKLESEMTL-----KEQHKEMKTLRLQYKTLREGNKQ--VOETS 1603  
QY 1163 LTLDTVTIGELHELADKAKLVSDHKPDADQ-----IKOLRQOQFTLR--EKRYE 1209  
Db 1604 LOLEVTYSKYDHV---KSKVLKDOQTFQEKORLLOVQELNKLQSQEKTIRSQOQKLK 1660  
QY 1210 SNPVKHYTDMGTTHNKALE-----ANYDAVKAFINAFKKEHGVNLTTR 1253  
Db 1661 QREGETHEEADKSHKRVLELESOLEQOQTAVEHYKAOEMKAKVHYDAKKKQONQELSSELO 1720  
QY 1254 TVLESQ-----GSAELAKKLKNTLLSLDGSMSFSRSYSGGVSTVFVPTLSK 1301  
Db 1721 SHIKOOEHLKSENADLKAESQHLKELHLSLQSKVEQN-----CKNLSN 1766  
QY 1302 KVPVPVPGAGITLDRAYNLNLSFRTSGGLNVSFGRGGVSGNIMVAT-----GHDVMPY 1355  
Db 1767 RV-----RSLEASWSTLTQOL-----RD---PCKFQLATAFHEESGHFCAPR 1805





QY 1554 NRTSOSISLEKRAEPTVSTNDISELTSTLTKGHFKDS- - - - -TTKMLAALKELDKAPBOL 1610  
Db 1165 SONGQ- - - - -NTITGLSNTLANVTNDKGSVRTTEQGNIIKDEKTRAASIV 1210  
QY 1611 HIQQHFS- - - - -AKDVGDERYEAVRNKLVQROAASHMELGSASHSTYNNLS 1664  
Db 1211 DVLSAGFNLQNGEAVDF- - - - -STYDTV- - - - -NFADGN- - - - -ATTAKVTYDDTS 1253  
QY 1665 RIN- - - - -NDGIVELLHKHF- - - - -DAALPAS- - - - - 1686  
Db 1254 KTSKVVYDVNVDDTTIEVKDKKLVGKVTTLTSTGTGANKFALSNOATGDALVKASDIYAH 1313  
QY 1687- - - - -SAKRLGEMMN- - - - -DPALKDIIKQLQSTPFSASVSMELKDLRQTEKA 1733  
Db 1314 LNTLSGDIQTAKGASQANNSAGYVDADGNKVIYDSTDNKYQAK- - - - -NDGTVDKTEV 1368  
QY 1734 ILDKVGREEV- - - - -GVLFQDRNNLRVKSVSQSV- - - - -SKSEGFNTP-ALLLGTSSAAMS 1787  
Db 1369 AKDKLVAQAOTPDGTLAQ- - - - -MNVKSVINKEQVNDANKQGINEDNAFVKGLEKRAASN 1424  
QY 1788 MERN- - - - -IGTIN- - - - -FKYGODQNTPRR- - - - -FTIEGGIAQANPOVASALTD 1828  
Db 1425 KTKNAAVTVGDNLNAVAPLTFAGDTGTAKKLGETLTIKGGQTDN- - - - -KLTD 1475

RESULT 2  
US-60-248-505-909  
; Sequence 909, Application US/60248505  
; GENERAL INFORMATION:  
; APPLICANT: Beasley, Ellen  
; TITLE OF INVENTION: ISOLATED HUMAN G-PROTEIN COUPLED  
; TITLE OF INVENTION: RECEPTORS, NUCLEIC ACID MOLECULES ENCODING HUMAN GPCR  
; TITLE OF INVENTION: PROTEINS, AND USES THEREOF  
; FILE REFERENCE: C1000918  
; CURRENT APPLICATION NUMBER: US/60/248, 505  
; CURRENT FILING DATE: 2000-11-15  
; NUMBER OF SEQ ID NOS: 1998  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 909  
; LENGTH: 3170  
; TYPE: PRT  
; ORGANISM: Human  
US-60-248-505-909

Query Match 2.1%; Score 198.5; DB 6; Length 3170;  
Best Local Similarity 18.5%; Pred. No. 0.0063;  
Matches 422; Conservative 289; Mismatches 778; Indels 791; Gaps 101;

QY 4 KSLGTEHKAHVTAAHNPVGH- - - - -GVALQGGSSSSSSPQNAASAAAEKGNRGMPIRHQPS 61  
Db 807 KDVGSEK- - - - -HVRPOIP-GRKWGGFLGRSKWDPQNKDK- - - - -EGVLLSKAEKTPGPQ 859  
QY 62 TAADGISAHQKQKPSLRCLGT- - - - -KKFSRPAQOGPQGTTHSK- - - - - 103  
Db 860 T- - - - -OMEKTSQVQELGDLRMRGKAGELRSTTGKAGSDGLDGRHAKKPPCK 910  
QY 104- - - - -GATLRDLARDDG- - - - -ETQHEAAAPDAARLTR 132  
Db 911 APDRWYEAQVLAQKDGFTLATVLPDEGTADLPAGRVRLCIDADKTITEVDEEHVR 970  
QY 133 SGGVKKRNMDDMAGRPMVKGGS- - - - -CEDKVPTQQRHOLNFMGM 174  
Db 971 ANPELDOVEDLASLISVNESSVLNTLQRYKAQLLHTCTGPDLLVLPQPGSPVPSACKP 1030  
QY 175 RQTMLSKMAHPASANAGRLQHSPPHIP- - - - -SHEIKEEPVSGTSKAT 220  
Db 1031 KASELSLTPRDSGSL- - - - -EIPGRDIFIDPSNOVLIGHPLGRHOMLGSAKRC 1081  
QY 221 TAHADRVETIAQEDDDSEFQQLHQORLARERENPPQPK- - - - -LGVATPISARQPKLTAVA 277  
Db 1082 RHWLDQPLIQ- - - - -MATRWEHRAVPKGRDGLPAHIGSMAQRAYWALL 1127

QY 686- - - - -HIGSLALQEG- - - - -KLHYFDOLTGWGTGAESDCKQL- - - - -KKGLDGAAYLLKQGE 732  
Db 203 ATSSSLGALAGAGTRAOQLGQSSVVTQSDNNSRPATPTQALDPKQFQATNNK 262  
QY 733 VKRLNINOSTSIKHGTENFVSLPHVRNKPPEGDALQGLNKDDK- - - - -AQAMAVIGVKNYL 789  
Db 263 AGPLSI- - - - -GNSIKRKIINVA- - - - -GVNKTDAVNVAQLEAVKWKAKER 305  
QY 790 ALTEKGRTSRQIKPGTQO- - - - -LERPAOTLSREGISGELKD- - - - -IHVDHKONLYALT 840  
Db 306 RITFGDDNDSTVKITGLONTLITKGAETNALTDNNGIVGVKEADNSGLKVLAKTLNNLT 365  
QY 841 HEGEVHFQPREAWONGAESSSWHKAL- - - - -PQSESK- - - - -LKSLSMSHE 882  
Db 366 EVNTTLNATTVKVGSSSTTAELLSDSLTFPTQPTGSGOSTSKTVYGVNGVKFTNNAET 425  
QY 883 HKPIAT- - - - -FEDG- - - - -SOHLKAG- - - - -GWHAYAAAPERGPLA 914  
Db 426 TAAIGTTRITRDKIGFARDGVDEKQAPYLDKKQKLVGSVAITIDNGIDA- - - - -GNKIKISNLA 484  
QY 915 VGTSGOTVFNRLMOGVKVIPIGSLTV- - - - -KLSAQTGGMT- - - - -GAE 956  
Db 485 KGSANDAVTIEQLKAAKPTLNAAGISVTPTEISYDAKSGNVTAPYINIGVKTIELNSD 544  
QY 957 GRKVSXKFSERIRAYAFNPTMSTPRPI- - - - - 983  
Db 545 G- - - - -TSKKS- - - - -VKGSGTNSLNTAEHLASLYNEVNRRTADSALOSFTVKEEDDDANAITV 600  
QY 984- - - - -KNAAYATQHGQWREGKPLVEMOGALIKQL- - - - -DAHNVRHNA 1023  
Db 601 AKDTTKNAGAVSILKLGKNGNLVATFKDGTVPFGLSODSLITGKSTLNDGLTVKDTN 660  
QY 1024 POPDLOS- - - - -KLETLDLGEHGAELLNDMKRFELEQSATRSVTVLQHQGVLSKNGEIN 1080  
Db 661 EQIOVGANGIKFTNNGSNPGTGIANTARITRDKIGPAGS- - - - -DGAVDTNKPYL 711  
QY 1081 SEFKPSGKALVGFVNNRSGDLSKLSQAVHATPPSAESKLSQMLGHFVSAGVDMHQ 1140  
Db 712 DQDKLOVGNVKTITNGINAGKAIT- - - - -GLSPTLPSI- - - - -ADQSSR 751  
QY 1141 KGIPLGRQDPNDKTKALSRILDT- - - - - 1173  
Db 752- - - - -NIELGNTIQDKDSNAASINDILNTGFLNKNNNPIDFVSTYDVIDFANGNATTATVT 809  
QY 1174 HELADK- - - - -AKLVSD- - - - -HRPDADQIKOLROQFDTLREKRYESNPVKHYTDMGFTHN 1224  
Db 810 HDTANKTSKVYDVNVDDTTIHLTGTDDNKKLVGKVTTKLNTSANGN- - - - -TATNFVN 864  
QY 1225 KALE- - - - -ANYDAVKAFINAFKKEHGVNLTTRTVLESQGSAAELAKKALNTLLSLDSGSSM 1281  
Db 865 SSDEDALVNAKDIAENLTLAKEIHTTKGTADTALQ- - - - -FTVKKVDENNADANAIT 920  
QY 1282 SFRSYGGGVSTVPTLTKKVPVPIPGAGITLDRAYNLSFS- - - - -RTSGLANV- - - - -SFGRDG 1338  
Db 921 VGQNNANNQVNTL- - - - -TLKGE- - - - -NGLNIKTDKNTVTFGINTTSLGKAGKTLNDG 971  
QY 1339 GVS- - - - -GNIMWATGHDVMPYMTGKKTSGNASDWLSAKIKISPDLRIGAAVSGTLOOT 1393  
Db 972 GLSINKPTGSIQVAGDVKFA- - - - -KVNNGVVGAGIDGT- - - - - 1010  
QY 1394 LQNSLKFPLKLEDELPGFTHGTLTPAE- - - - -LLOKGTIEHQMKQSGKLTFFSVDTSANLIDL 1451  
Db 1011- - - - -RITRDEI- - - - -GFTG- - - - -TNGSLDKSKPHLSKDGIN- - - - -AGGKKIT- - - - -NI 1048  
QY 1452 RAG- - - - -INLNE- - - - -DQSKPQNGVATARYSAGLSANLAAGSRSTTSQGFSTTSASNNRP 1506  
Db 1049 QSGEIAQNSHDAVGTGCKIYDLKTELENKLSSTAKTAQNSLHEFSVADEQGNFTVSNPYS 1108  
QY 1507 TF- - - - -LNGVGAGANITAAIGVAHSSTHECKPVGIIPTFTSTVNSAALD 1553  
Db 1109 SYDTSKTSVDITFAGENGITTKVN- - - - -KGVVRVIGIDQTKGLTTPKLTGVNNGNGKIVID 1164



QY 278 ESVLEGTDTTQSPKQSMKLGSGAGVTPPLAVTLDDKGLQLOLAPDNPALNTLLKQTLGKD 337  
Db 1128 -----NORRDSIIVGRSGAGKT-----TCCQOVL-- 1151  
QY 338 TQHYLAHASSDGSQHLHLLDKHGLFDIKSTATSYSLVHNSHPCEIKGKLAQAGTGSVSV 397  
Db 1152 -EHLVGMAGSDVG--RVSVEKIRATFVLRFAFGSVMAHS-----RSATRFWSWMSL 1200  
QY 398 D-GKSKISLGSSTOSHINKTMSLQOPGAHRSLTGTIWOHPAGARPQGESIRLHDDKIHI 456  
Db 1201 DFNATGRI-----TAAQLOTMILLEKSKV-----ARQPEGESNLFVSQMLA 1241  
QY 457 -----LHPELGVWQSDAD-----KQTHSOLSRQADGKLYALKDNRTLONLSDNKSS 501  
Db 1242 GLDLRLTELNLHOMADSSSGFVMSKPEDKQKAAAFALQLOG-----AMEMIGITISE 1296  
QY 502 EKLVDKI-----KSYSDVORG-----QVAILTDPGRHKMSIM 534  
Db 1297 QRAVWRVLAAYHLGAAGACKVGRKQFMRFEWANYAAEALGCEYEELNTATFKHHLRQII 1356  
QY 535 PSLDASP-----ESHISLSLHFADAHQGLLHGKSELPQSVSA-----TSHGR 576  
Db 1357 QOMTEGSPRWGLEDETSSGLKMTGVDCVEGMASGO-ELFAAVVSLINRSFSSHLSMAS 1415  
QY 577 LVVADSEGLFSAATPKQGDGNELMKMPQHALDEHFGHDHQSIGFFH----- 625  
Db 1416 IMVVDSPG-----FQNPRIHQGDRAATFEELCHNYAHERLQLLFYQRTFYSTLORYO 1467  
QY 626 -----DDHCQNALVKNFRQOHACPLGNDHGFHGWNL-----TDALV 664  
Db 1468 EEGVPVQFDLPSPGTTAVVDQN-PSQVRUPAGGAGDRLFWLDEEVHVEGSDSVV 1526  
QY 665 ID-----NOLGLHHTNPPEHPHIDMGHLGSLAL-----QEGKL 697  
Db 1527 LERLCAAFKKGAGTEGTFACESHKGQOGLGSLTNEPRH---SQSGSSALRTCEQPLQC 1583  
QY 698 HYFQDLTKGW-----TGAESDCKQLKGLDGAAYLLKDGVEKRLINQSTSSIK----- 746  
Db 1584 EIFHQL--GNDPVRYDLTGLWHRAPKNLSALDAPQVLOOSKREELRSFQARAKLPPVCR 1641  
QY 747 -----HGTEN-----VFSLPVRNKRPEPG-----DALQGLNKDDKAQAMAV 782  
Db 1642 AVAGLEGTSOALORSRMVRRTFASSLAARVRKAPCSQIKLQMDALTSNKRSLRHF 1701  
QY 783 IGVNKYLALTEKDIRFOIKP-----GTQOLERPAAOTLSREGISGELKDIH 829  
Db 1702 LVPNPVV-----ESRSGQSPPPPQCRDKPGAGPLALDIPALRVQLAGFHIALR-LH 1754  
QY 830 -----VDHKQNLIALTHEGEVPHOB-REAWONGAESWHKLLALPQSESKLSLMSHEHK 884  
Db 1755 RTGYADH-MGLTRFRQFQVLDAPLLKLLASTSEGIDERKAV-----EELLETLDL--EKK 1807  
QY 885 PIATFEDGSHQHLKAGWHAYAAPERGLAVGTSGSTVFNRMLMQGVKGVIPGSGLTVK 944  
Db 1808 AVAV--GHSQVFLKAG-----VISRL-EKQREKLVQSOS-IVLF 1841  
QY 945 LSAQGTGMTGAEGRKVSSKFSERIRAYFNPTMTSPPIKNAAYATOHQWO----- 995  
Db 1842 QRAKGFSLRQEPFKLIR---RLAQCIQKNVAVFLAVKDWPM-----WOLLGSLQPLL 1893  
QY 996 -----GREGKLPYEMOGALIKOLD-AHNVRHNAPOPD--LQSKLETLDLGEHGAELNDM 1048  
Db 1894 SATIGTEQLRAKEEELTLRLKLEKSEKRLNELRQNTDLLESKI-----ADITSDL 1944  
QY 1049 --KFRFDE-----LEOS-----ATRSVTVL-QHQGVKLSNGEINSEFPKSPGKALVQSFN 1096  
Db 1945 ADERFKGVACQVLESARLQAFREVOELKSEHQVQKGLGVDYKNKOLEEAAQKIQLNDLE 2004  
QY 1097 VNRSQ-----ODLSKSLQOAVHA---TPPSAESKLSQMLGHFVSA---GV 1135  
Db 2005 RNPTGADWQMRFCDAWENEFRLRQOCEERLDSLTARKELEQKLGELOSAYDGAK 2064  
QY 1136 DMSHQKE-----IPLGRQRPNDKNTALTKSRLLI-----D 1166

Db 2065 KWAHQLRKCRHHHTUCLEDTCVLLENQOOSRNHELEKKOKKFDLQLAQALGESVFEKGLRE 2124  
QY 1167 TVT-----IGELH-----ELADKAKLYSDHKPD----- 1189  
Db 2125 KVTQENTSVRWELGLOQOOLKQKEQASOLKQOQVEMLODKHRELLGSPSLGENCACVAGLKE 2184  
QY 1190 -----ADQIKQRQOF-----DTLREKRY 1208  
Db 2185 RLWKLESSALEOOKIQSOQENTIKOLEURLORFELEITERMKQMHQKQDREDOEBELEDVRO 2244  
QY 1209 ESNPKVHYTDMGF-----THNKA-LEA-----NYDAVKAFINAFKKEH 1245  
Db 2245 SCQKRLHOLEMOLQEQEYEBEQMVLEHEKQLEGLEGITGLDQIGHRDEFEKRLRDLRTH 2304  
QY 1246 ---HCVNLTRTVLSESQSA---ELAK---KLKNTLLSLDSGESMSFSRSYGGGVSTVFPV 1297  
Db 2305 ALLSDVQLLLGTMDGKTSVSKEELEKQVHSQSSTLLALKKGTSSGMWGLYNKAVVTLVPQ 2364  
QY 1298 TLSK-----KVPVPVIP--CAGITLDRAYNLSFSRTSGGLNVSGRDGGVSGNIMVAT 1348  
Db 2365 VISSYRGSCQVPPVLPVDTVGTWLEQ----- 2393  
QY 1349 GHDVWPYMGKTSAGNASDMLSAAKHKISPDLRIGAAVSGTLOGTLONSLKFKLTEDELP 1408  
Db 2394 -----SEAKCEALKTKOKVLTADLE-----SMHSELENTRKNSLVE-- 2431  
QY 1409 GFHGLTHGTLTPAELLQKQIEHQMKQSKLTFSDVTSANLDRAGININEDGSKPNGVT 1468  
Db 2432 ---OLYRLOPEKADLLKRIEDQ-----DDLNELMQKHKDLI 2465  
QY 1469 ARYSAGL---SASANLAAGSRERSTTSGQFGSTTS-----ASNRPPTFLNGVGAGANLT 1519  
Db 2466 AQSAADTCQIQELQLOLEEAKEKHKLQEQAGSCCTCLPCGIRDNCRPIDFSITYSCRWL 2525  
QY 1520 AALCVAHSSTHEGKPVGIFPAFTSNVSAALA-LDNRTSQSISLEIKRAEPTVSDISEL 1578  
Db 2526 CA-----SSTWNSP-WYEP---SSAGRAKAVICDLENKT-EFKQVQIKRFEVPSIQAPG 2576  
QY 1579 TSTLGHFKDSATTKMLAALKELDDAKPAEOLHILQOHFSKADKVVGDYERYAVR-NLAKKL 1637  
Db 2577 TTVLVIRLRDS---LIKMGEEELSQAATSES---QRESSQ---YYQRRLEELKADMBEL 2626  
QY 1638 VIPOQADSHSMELG-----SASHSTYNNLSRINNDGIVELLHKKHFDAAALPASSAKRL 1691  
Db 2627 VQREAEASRCMELEKYVEELAAVROTLDLE-----TSIRRI 2665  
QY 1692 GEMMNDPALKDIKLOLQSTPFSASVSMEIKDCLREOTEKAILDGKVGREVGVLFODR 1751  
Db 2666 AD-----LOALEEVASDSDESQVAVDCG---SSGRKEIIRGHKSRVPVLKMIQIR 2716  
QY 1752 NMLRVKSVSVSOSVSGSEGTNTPALLIGT-----SNSAAMSMERNIGTINFKYQDQNT 1805  
Db 2717 DNVSILS---SQPEGLSQWLSCTLSLATDMRTPTSPQSATSSILSPRINEEAGDERT 2773

## RESULT 3

US-09-739-449-10549  
; Sequence 10549, Application US/09739449  
; GENERAL INFORMATION:  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof  
; FILE REFERENCE: 38-10(15490)C  
; CURRENT APPLICATION NUMBER: US/09/739,449  
; CURRENT FILING DATE: 2000-12-19  
; PRIOR FILING DATE: 2000-02-23  
; NUMBER OF SEQ ID NOS: 13351  
; SEQ ID NO 10549  
; LENGTH: 2031  
; TYPE: PRF  
; ORGANISM: Agrobacterium tumefaciens

Query Match	2.1%	Score 196.5	DB 5	Length 20311
Best Local Similarity	18.9%	Pred. No. 0.0038		
Matches 337	Conservative 249	Mismatches 714	Indels 485	Caps 73
QY	325	ALNTLLKOTLGKDTQHYLAHHASDGSQHLLLDNKKHLFDIKSTATSYSVLHNSHPGEIK	384	
DB	121	AMNDGIERTIARATELETLVHSEYNALERSYADNELKRSVLQBELTARDAIVNAHERIR	180	
QY	385	GKLAQAAGTGSVDGSKKISLGSSTGSHNKTMLSPQCE--AHRSLTGTIWHQHPAGARP	442	
DB	181	SSIV-----GAQEQIKEELSIYGEELSMRIATTG-----EAFASMI	216	
QY	443	QGESIRLHDDKIHILHPELGVWQSAKDTHSOLSRQADGKLYALKDNTLQNLSDNKSS-	501	
DB	217	DTRSAALE-----KSRASTEAMGSLTAAKTENLLQALNSGGSTI	256	
QY	502	-----EKLVDKTKSYVDORGVAAILDTPGRHKMSIMPS-LDASPES-----H	544	
DB	257	SNEFDMRLHUNITS-TLDRGEVLL-----ERFAIHASTLSDGSVESLSALEERTQLN	308	
QY	545	ISLSLHFADAHOGLLHGKSELEAQVAISHGRVLVYADSEGKLFSAAPKQGG-----	597	
DB	309	ETLSARSLELNRIGRQGVIGGSLDVLDKLSTLEEKGLSFROSLOSTADDAIMDLDL	368	
QY	598	-----NELKMKCA--MPQHALDEHFHDHQISGFFHDDHGQLNALYKNFNROQHACPLGN-	649	
DB	369	RSGLYEERQATVGVNSAFDEHVA--QFASAFDQAGSLDSKLMESLARINETVAGGS	425	
QY	650	-----DHOQHPGWNLTDAVLIDNOGLHHTNPPEHEILDMG-----HLGSLALOGKHLHYFD	701	
DB	426	EALDITLTSGLERIGTMDQSLATALGTQOEMLENALESRTQAFSAIGORTAEITD	485	
QY	702	QLTKWGTGAESDCOKLKGLDGAAYLLKD--GEVKRLINTNQSTSSIKHGTENVFSLPHVR	759	
DB	486	AFTNSHAKIDTVLAERSNALFGALSASQDRFDEALASRLAITSVSGTAEHLAAM----	541	
QY	760	NKPEPGDALOGLNKDKQAQAMAVIGNYKYLATEKGDIRSFQKPGTQOOLE-RPAOTLSR	818	
DB	542	-----LDERAAINSVADVERLTETLETRAATITGAVSGIEDRISDTL--	586	
QY	819	EGISGELKDIDVHDKNLYALTHEGEVFHOPREAWONGAESSSHKKLALPQSESKLSLD	878	
DB	587	ESRTAALHDV-VSGAESRIADTLDG-----RTAALSSAISGVBEERTA-DTMSRTLSLL	638	
QY	879	MSHEKPIATFEDGSQHOL--KAGGWHAAYAAPERGLAVGTSGTSQTVFNRLMQCVKQVI	936	
DB	639	MT-----FANVEERSETLNRFSALTGCIYASAEKTAGALDSRTATFGDVVAGAETRIA	693	
QY	937	PG-SGTVTKLSAQTGGMTGAERKVSCKFSERTIRAYAFNPTMTPTPRIKNAYATQHGWQ	995	
DB	694	ETLDGRTAALNAV--VSGAEERIADLDS--RTMALDITFGSAEKKIAEALDTRTAAL	747	
QY	996	GREGLKPLYEMOGALIKOLDA-HNVRHNAPO--PDLOSKLETLDLGHGAELLNDMKRF	1051	
DB	748	GELVASAETRIAGALDSRTSLKTVVSGAEERITDVLDSRTMALDMSFGVE-----EKI	802	
QY	1052	RDELE-QSATRSVTVLGOH-----OGVLKSNCEINSEKFPSPCKAL-----	1091	
DB	803	TDILDGRTAALKSAVAGVEDRIAGALDSRTAALSGIVSGAEERIAEALDSRTLADMTIS	862	
QY	1092	-----VQSFVNRSQGDLSKS-LQQAQVHATPPSAESKLSQSMIGHFVSAGVDMSHQGEI	1144	
DB	863	GVEERIAEAMDARASSLSLAAGVGQRLEATAFTLENALAS-----GHERLET	910	
QY	1145	PLGRQRDPNDKTALTKSLIIDLTVTIGIEHELADKAKLVSDHKPPADQIKQLRQOQFDLRL	1204	
DB	911	MLGSAERIAIAGSLERNISGLIPOSV-----	935	
QY	1205	EKRYESNPVHYTDMGTFTHNKALEANYDAVKAFINAFKKEHGVNLTRTVLESQGSAPL	1264	



271 ---ESPTRSNAPKPLSPKDVVASPKLPERESERSRSQSQLKOTDTSSEGSPPRENPREABG 327  
256 ---PPKLGWATPISARFQKLTAVASV-LEGDTTQSP-LKPQSMKGGAGVTPPLAVTL 311  
328 ELPSGGGPAAPPDAELSPRWSSAAVAVQCHENDSPOLEP---LEAGEPEF-PDATTT 383  
312 DKGLKOLAPDN---PPALNTLLKQTLGKDTQHYLAHHASDGSQHLHLLDNKGHLF 363  
384 AK-QLHSSFGSSROENAEPTA---RSPGEDASPGAGH---EQEAFGLVGRG--- 427  
364 DIKSTATSYVLHNSHPG---ELKGLAQAGTSVSDGKSGKISLGSQTQSHNKTML 418  
428 ---APGSPTQERPAAGLGEAPNAAPSVAEAGSGSGPRNARSQPPKAS 472  
419 SOPGE---AH---RSLLTGIWQHPAGAARP---QGESIRLHDDKIHL 457  
473 DLPGEPBAGCAHTASSAQADCTARKGHAHPAKVLTLDIYLSKTEGAQV---DEPVVI 528  
458 HP---ELGYWQSAKDTQSLSRQ---ADGKL---YALKDNRTLON- 494  
529 TPRAEDCGDWDMEKRSRGRSGRSQKSTDPGADAEAPESAAARDADAVDEAVAPNA 588  
495 LSDNKSSEKLVKIKSYVDORGVAILTDPGRHKMSIMPSLSDASPEHSISLSLHPADA 554  
589 ASDNASAEK---KVKSPRAALDGVASA---ASPESKPS---PG 623  
555 HQGLLHCKSE---LEAGSVAISHGRV--- 578  
624 TKQLRGESDRSKOPPPASSPTKRKGRSRALEAVPAPGSPGRAPAKESPVKRVPDPSVP 683  
579 ---VADSEGLKLSAATPKQDGNELKKA---MPQH---ALDEHF---GHDHQ 619  
684 TKGTAAESGEAARATPR---ELPVKSSLLPEIKPEHKGRLPLNHNFGRAEGRSRE 738  
620 IS---GFFHDD---HGQNALVKNKFNROHQACPL 647  
739 LGRAAGAPGASDADGLKPRNHFVGRSTVTTKVTLPAKPKHVELNLKTPKNLDS---L 793  
648 GNDHQ---FPHGWNLTDALVIDNOLGHLHTNPEPEILDMHGLSLA---LOEGLK 697  
794 GNEHNPSQPVHKGNTATKISLFENK---RTNSPRTDTRGORTNTPASSKTFVGRAKL 849  
698 HYFDLTKWGTGAESDCKQLKGLDGAAYLALDKGVEKRLNINQSTSTKHGTENVFSLPH 757  
850 N---LAKKAKEMEQPEKKVMPNSPQNGVLVKEAIE---TKVTVSEEEI---LPA 895  
758 VRNKEPFGDALQGL---NKDDKAQ---AMAVIGVNKYLALTEKGDIRSQ 801  
896 TRGMNGDSSENOALGPPQNDKADVQTDAGCLSEPVASALIPVKDH-KLLEKEDSEAD 954  
802 IKPGTQQLERPAQTLRSREGISGELKDI---HVDHKONLYALTHEGEVFFHQPREAW---QNG 856  
955 SK---SLVLENTVDTAQDIPTVDTKD---LPTTAMPKPQHTFSDSQSP 997  
857 AESSSWHKLAL---POSESKLSLMSHEHK---PIATFEDGSQHQ 896  
998 AESSPGSLSLSAPAGDPVDPKTCVQSPITSSPPCTDLKSYENHKGCVLPVSRQNNKEMPL 1057  
897 LKAGWHAYAP---ERGLAVGTSQGTVPENL---MQGVKGVKIPVSGSLTV 943  
1058 LELGG---ETTPPLTERSEPAVGSECPVSRVLQVRSFVLPVESTQDVSSQVITES--- 1110  
944 KLSAQGTGMTAGRGKVSSEKFSERIRAYAFNPT---MSTPRPKNAAYATOHGQGREGLK 1001  
1111 ---SEREVQL---PTCHSNEPEVSVASCA---P 1136  
1002 PIYEMOGALIKOLDAHNVHNAPOPDLOSKLETLDLGEHGAELLNDMKRFRDEQSATR 1061  
1137 POEEVLGNEHSHCTAEPLAAKSGPQ---VIPPASEK 1168  
1062 SVTVLGOHGVKLSNGEINSEFPKSPGKALVOSFVNRSGDLSKLOQAVHATPPSAES 1121  
1169 TLPQAOQSG---SRTPPLMAESSPT---NSPSSGNHL---ATPQRPDQTVTNGQD 1214

QY 1122 KLOSMGLGHFYSAGVDMSHQKEIPLGRORDPNDKTALTAKSRLLDVTVTIGELHELADKAK 1181  
Db 1215 SPASLLN--ISAGSDDS----- 1229  
QY 1182 LVSDHKPDAQIIKQLROQFDTL-----REKRYESNPVKHYTDMGFTTHNKALEANYD-A 1233  
Db 1230 -VFDSSDMKEFTETIIKQMSAVCMKPKRKKARMPNSPAPHA-MPIIHEDHLEKVFDPK 1287  
QY 1234 VKAFINAFKE-----HHGVNLTTRTVLE-SQSAELAKLKNLTLLSLDSDGESMSF 1283  
Db 1288 VFTFGLGKKKESQEPMSPALHMQNLDTKSLRKPRAAEQSVLFK-----SL 1335  
QY 1284 SRSYGGGVSTVFVPTLSKKVPPVPIPGAGITLDRAYNLSFSTSGGLNVSVFQDGVSN 1343  
Db 1336 HTWNGNSEPLVMEINDKENDV-----TNGGIKRSRLEKSALFSS 1377  
QY 1344 IMVATGHDVM--PYMTGKTSAGNASDMLSAKHKISPDLRIGAAYSGTLOGTQNSLKFK 1401  
Db 1378 LLSSLPQDKIFSPSVTSYNT-----MTATSTSONGSLSSQSVSQ 1417  
QY 1402 LTEDLPGFHGLTHGTLTPAELLQ-----KGIEHOMKOGSKLTFSDVDTSA 1447  
Db 1418 PTTEGAPPCGLNKEQSNLLPONSLLKVFNFNSSTSHSLKSPSHMEKYPQK-----EKTKE 1473  
QY 1448 NLDLRAGINL-----NEDGSKPNGVTARVSAGLSASAN----- 1480  
Db 1474 DLDSRSLNHLPETKFSLSKLKNDMEKANHIESVIKSNLPNCANSDDTDFMGLFKSSRYD 1533  
QY 1481 --LAAGSRERSTTSGQFSGTTSASNNRP 1506  
Db 1534 PSISFSGMSLSDTMTLRGVSQVKNLNRPP 1561  
RESULT 6  
PCT-US01-04098A-3914  
; Sequence 3914, Application PC/TUS0104098A  
; GENERAL INFORMATION:  
; APPLICANT: Hysed, Inc.  
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides  
; FILE REFERENCE: 21272-029  
; CURRENT APPLICATION NUMBER: PCT/US01/04098A  
; CURRENT FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: Not Yet Assigned  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: 09/728,422  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: 09/693,325  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 09/663,561  
; PRIOR FILING DATE: 2000-09-15  
; PRIOR APPLICATION NUMBER: 09/654,936  
; PRIOR FILING DATE: 2000-09-01  
; PRIOR APPLICATION NUMBER: 09/620,325  
; PRIOR FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 09/598,075  
; PRIOR FILING DATE: 2000-06-20  
; PRIOR APPLICATION NUMBER: 09/560,875  
; PRIOR FILING DATE: 2000-04-27  
; PRIOR APPLICATION NUMBER: 09/496,914  
; PRIOR FILING DATE: 2000-02-03  
; NUMBER OF SEQ ID NOS: 3960  
; SOFTWARE: Custom  
; SEQ ID NO 3914  
; LENGTH: 2515  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
PCT-US01-04098A-3914

Query Match 2.0%; Score 193.5; DB 1; Length 2515;  
Best Local Similarity 17.9%; Pred. No. 0.0086;  
Matches 377; Conservative 286; Mismatches 806; Indels 633; Gaps 94;

QY	28	LOOGSSSSPQNAASLAA-----EGKNRGKMPRIHOPSTAADGI--SAAHQOKKSFSLR	80
Db	268	LKHVHAKSEPKPARRLSLHVVDENKNKESKIEREHKRTSTPVMGCVQSETDTRDYK	327
QY	81	GCLGCTKKFSRAPOGQPTTHSKATLRLDLIARDDDGTOHEAAAPDAARLTGSGGVKRRN	140
Db	328	ROVPERSEICTEPOKQKSTLANE-----KHLKDDSETPHLKS-----LLKKE	370
QY	141	MDDMAGRPVYKGGSGEDKVPYQOK-----RHOLNFGQMR-----OTMLSKMAHPAS	187
Db	371	VKSSEKPEREKTPSEDKLSVKHKYKDCMKHKTGDETLHSEKGLKVKEENIQOSQOTK	430
QY	188	ANAGDRLQHSPPHPHGPSSHEHKEPEPVGSTSKATTAHADRV--EIAQEDDDSEFQOL-----	241
Db	431	LSDDKTKERKSH-----RNERKLSVLGDKGQVPSEYIKTIDENVRKENNKKERRLSAET	486
QY	242	---HOORLARENPPOPPLGLGVATPTISARFOPKLTAVAESVLE--GTDTIT--QSPLPK	293
Db	487	KAEHKSR--RSSDSKIQKDSLG-----SKHGITLQRSESYSEDKCOMDSTNMDNLKP	539
QY	294	OSML-----KSGSAGVTPLAYLTDLKGKLLQLAPDNPP-----AL	326
Db	540	EEVVHKEKRRKTSLLBEKLVKSKSTQGRQVKVVELOEGATKQATTPKPKDEKNTTEE	599
QY	327	NTLLKQTLGK-----DFOHYLAHHASSDGSOH--LLLDNKGHLFDIAKSTA	369
Db	600	NDSEKQKSKVEDKPFEECTVEPVLETATSSAHSITQKSSHRAKLPPLAKBKYSKDKDSTS	659
QY	370	T-----SYSVLHNSHPGEIKGLAQAGCTGVSVDGSKGSKISLGSGETQSHNK---	415
Db	660	TLERKLSGDHKSRSRLKHSK--DIKKK-----DENKSDDKGK-----EVDSSHEKARG	707
QY	416	-----TMLSQPEAHRSLLTGIWQHPAGAA--RPQ--GESIRLUHDD	452
Db	708	NSSLMEKKLSRRLCENRRGSLSQEMAKGEEKLAANTLSTPGSSSLORPKPKSGDMTLIPEQ	767
QY	453	KTHILHPGLVWQSDAKDTHSOLSRQADGKLYALKDNRTLQNLSDNKSSEKLVDKISYS	512
Db	768	EPMEIDSEPGV-----ENVFEVSKTQDMNRNNSHQDIDSENNKQKTSATVQKDELRTCT	821
QY	513	VDRQGVAILTDPGRHKMSIMPSLDASPESH-----LSLSLHFADA-----HQGL	558
Db	822	ADSKATAPAY--KPGRG-----TGYNSENKKAHHRSTLTTKMHITOSAVSKMNPKEKEPI	874
QY	559	LHGKSELBAQSAVISHGRLVWADSEKGLFSAAIPKQGGDGNELKMKAMPQHALDEHFGH--	616
Db	875	HRGTTEYNIDSETV--HRMLLSAPSE-----NDRVQKNLKTAAEEHVAQGD	919
QY	617	---DQISGFFPDHGGOLNALVKNNFRQOH---ACPLGN-----DQHPHP	655
Db	920	ATLEHSTN---LDSPSLSSTVTVPLRESYDPDVTPLFDKRTVLECGSTASTSPAGHSALP	976
QY	656	GWNLIT-----DALVIDNOLGLHHTN-----PEPHE--ILDMCHGLGSLAQ	693
Db	977	NQSLTVRESEVLYKTSDSKEGGEFTVDTPKAKASITSKRHIFEAHQATLLD--GKQGVIMP	1035
QY	694	EG--KLHYFDOLTKWGTGAESDCKOLKCLDCAAYLLKDCGEVKRLNINOSTS---SIKHGT	749
Db	1036	LGSKL-----TGVIVENENITK--EGG---LVDMAKKENDLNAEPNLKOTIKATV	1080
QY	750	ENVFSLPHVRNKPPEGDALOGLNKDDRAQAMAVIGVNKYALTEKDIRSFQIRPGTQOL	809
Db	1081	EN-----GKKDGIADVHVGLN-----T	1098
QY	810	ERPAQOT--LSREGISGELKDIHYD---HKONLYALTHEG-----EVFHO-----	848
Db	1099	EKYAETVKLKHKSPGKVKDLSIDVERNRNSENSEYDTSAGSGSAPSVLHRNQOTEDVATG	1158
QY	849	PRAWQNGAESSSWHKLLALPQSESKLSLDMSHEHKPIATFEDGSOQLKAG-----	900
Db	1159	PRAEKTSVATS-----TEGKDQVTLSPVKAGPATTTTSETQSEVALDPTCSIEA	1209

Qy	901	-----GMHAYAAPERGP	LAVGTSGSQT	VNRLMQGVKGI	PGSLGTVKL	SQAQTGMTG	954
Db	1210	DEGLIIGTHS----	RNNPLHVGAASE-	:	:	: :	:
Qy	955	AEGRKVSSKFSRTRAF	NPTMTPRIPKNA	YATOHGWOGREGL	KPLXYEMOGALIKOL	1014	
Db	1247	TGFAESETFLTSTFK-	:	:	:	:	:
Qy	1015	DAHNVNRINAPQDLO	SKLETILDGEHAEL	LNDMKRFDR-ELBQS	ATRSVTVLGHQHGVL	1073	
Db	1282	AVHAVKIEANVNSVT	EEXDDAVTSAGSE	BCKDGSIRSEIIV	EGITIFISEVESDGAVT	1341	
Qy	1074	KSNGEI-----NSE	KPKSPGKALVSFVN	RSGDLSLQAOAHATP	PSAESKLQSMML	1127	
Db	1342	SAGTEIRAGISSEEV	DGSG-----NMNR	MGP--KKETEGVTCT	CAGERS-----	1386	
Qy	1128	GHFVSAGVDMSHOKG	-----IPLGRDPN	DKTALTAKSRLLID	TVTIGELHELADKA	1180	
Db	1387	DNFVICSVTGAGPRE	RMVTGAGVLGDND	APPCTSASOBGDG	SVNDGTEGE--SAVTS	1444	
Qy	1181	KLYSDHKPADQIKOL	ROQFDTLREKRYE	SNPVKYHYDMGFTH	NKALEANYDAVKAFINA	1240	
Db	1445	GITEDCEGPA-----	-----SCTGSED	SSSEGFAISSSEENGES	--AMDST	1483	
Qy	1241	FKEHHGVNLTT-----	RTVLESOGSABELAK	KLKNTLLSDSGESM	SFRSYGGGVSTV	1294	
Db	1484	VAKEGYNPLVAAGP	CDDEGITVTGAKE	DEEDEGEDVVIST	CRGNFI-----CHASTC	1536	
Qy	1295	FVPTLSKKVPVIPAG	ITLDRAYNLSFRST	SGGLNVSFGRDGVG	SVNIMVATGHDVMP	1354	
Db	1537	-----TGLGESE	GVLCIESAEGDSQI	-----GTFVEH	FEAEAGAAIM-	1574	
Qy	1355	YMGKKTSGNADMWLS	AKKHITSPDLRIGA	AVSGTLOGTLONSL	KFKLTDELPCFIHCL	1414	
Db	1575	-----NANENNYD	MSGSTERKSGKDTC	ICSSAKGIVESSVT	SAVSGK--DEVTP	-VPGG	1624
Qy	1415	THGTLTPAELLQKGI	EHOMQCSKLTFSV	DTSANLDRAG-INLNED	CSKNPGVTARVSA	1473	
Db	1625	CEGPMTSAASDQ-----	SDSOLEKVEDTIT	STGLVGYSYDVLVS	GEVPECEVAHTSP	1676	
Qy	1474	GLSASANLAAGRSR	ITSGOFGSTTSASN	RPTFLNCVGAGANLT	AALGVAHASTHEGK	1533	
Db	1677	SEKEDIIT-SVENEE	CDGLMATASGD---IT	NONSLAGGKNQKVLI	IISTTTNDYT	1732	
Qy	1534	PVGFPAPTSTNVSA	ALAALDNRITSQSI	-----SLELKRAP	-VTSNDISELT-----	1579	
Db	1733	P-----QVSAITD	VEGGLSDALRTEEN	REGTRVTTTEEPAMP	SAVSGDDSQLTASREEK	1788	
Qy	1580	-----STLCKHKF	---DSATT-----KMJ	AALKE-----LDDAK	1605		
Db	1789	DECAMITSGEEFEL	PISATTIKCAESLP	QAAAAVEERATGP	VLISTADFEGMPSAP	1848	
Qy	1606	PAEHLILOOHFSAK	DYVGVDERVEAVRN	LKLVIROCAADSHMEL	GSASHSTYNNLSR	1665	
Db	1849	PEAESPLASTSKEEK	DECA-----LIST	STAEBCEASVGVV	VESENERAGTVME	1899	
Qy	1666	INNDDIVELLKHFD	AALPASSAKRKLCE	MMNNDPALKIIOQL	OSTPF-----SSASV	SME	1721
Db	1900	KDGSIIISTSSVE	-DCBGPVSSAVPOE	---GDPSV-----TP	AEMGMTAIMSTS	1946	
Qy	1722	LKQGLRQTEKAIL	LDKGKVGREEVGLF	QDRNNL---RVKSV	SVSQSVKSEGTNP	PALLL	1778
Db	1947	TSEGC-----EAVN	-----IGAVLQ	DEDRLTITRVEDLS	-----DAALI	1980	
Qy	1779	GTNSAAMSERNIGT	INFKYGDQONTPR	FTLEGIAQNPQVASAL	---TDLKK	EGLEM	1836
Db	1981	STSTACEMPISAI	-----DRHEENOL	TAD-----NPEG	NGDLSTATEVSKHKVPM	2025	
Qy	1837	KS	1838				
Db	2026	PS	2027				

Query Match	2.0%;	Score 193.5;	DB 1;	Length 2515;
Best Local Similarity	17.9%;	Pred. No. 0.0086;		
Matches 377;	Conservative 286;	Mismatches 806;	Indels 633;	Gaps 94;

Qy	28	LOQSSSSPQNAASLAA-----EGKNRGKMPRIHQPSTAADGI-SAAHQOKKSFSLR	80
Db	268	LKHVHAKEPSKPARLSESLHVVDENKNEKIEREHKRTSTPVMIGVQSEITDROVK	327
Qy	81	GCLGTHKFSRPAQOPGQTHSHKGATLRDLLARDCCETOHEAAAPDAARLTRSGGVKRRN	140
Db	328	QVRSSEICTEPEOKOKSTLKNE-----KHLKDDSETPHLKS-----LLKKE	370
Qy	141	MDMWAGRPVMYGGSGEDKVPYTOOK-----RHQLNFGQMR-----QTMLSKMAHPAS	187
Db	371	VKSSEKPEREPTSEDKLSVKHKYKDCMHKTGDETLHSSBKGLKVENIQKSOQTK	430
Qy	188	ANAGDRLOHSPPHIPGSHHEIKEEPPGVSTSKATTAAHDV-ETAAQDDDDSEFOOL-----	241
Db	431	LSSDDKTERKSKH-----RNERKLSVLGDKGKVPSEYIIKTDENVRKENNKKERRLSAET	486
Qy	242	---HQOBLARERENPOPPKLGWATPISARFQKFLTAAVESLVE---GTDTT--QSPLPK	293
Db	487	KAEHKSR--RSSDSKIQKDSL-----SKQHGITLQRRSESYSEDKDCMDSTWNDSNLPK	539
Qy	294	QSMLE-----KSGAGVTPVLAITLDKGLQIQLAPDNPP-----AL	326
Db	540	EEVYHKEKRRTKSLLEKLVLSKSKTQGGKQKVVELOEGATKQATTPKPOKEKNTEE	599
Qy	327	NTLLAKQTLGK-----DTQHYLAHHAASDGSOH-----LLLDNKGHLFDIKSTA	369
Db	600	NDSEKQKSKVEDKPFETGVEPYLETASSSAHSTQKSDSSHRAKLPLAKYKSDKOSTS	659
Qy	370	T-----SYSLVHNSHPGEIKGLQAAGTGSVSDGSKGKISLGSGTQSHNK-----	415
Db	660	TRLERKLSGDGHSKRSKLHSSK--DIKKK-----DENKSDDKDGK-----EVDSSHEKARG	707
Qy	416	-----TMLSQPGEAHRSLITGIWOHPAGAA--RPQ--GESIRLHDD	452

Db 1575 -----NANENNVDMSGTEKSGKDTDICSSAKGIVESSVTSVSGK-----DEVTP-VPGG 1624  
 QY 1415 THCTLPAAELLOKQIEHQKQGLTFSVDTISANLDRAG-INNEDGSKPNGVTARVSA 1473  
 Db 1625 CEGPMTSAADQ-----SDSOLEKVEDTITSTGLVGSYDLVLSGEVPECEVAHTSP 1676  
 QY 1474 GLSASANLAAGSRSTTSQGFQSTTSASNRRPTFLNGVGAGANLTAALGVAHSTHEGK 1533  
 Db 1677 SEKEDEIIT-SVENEEDCLMATTASGD---ITNQSLAGGKQGVLIISTTTNDYT 1732  
 QY 1534 PVCIFPAFTSTNVSAAALDNRTSQSI-----SLELKRAEP-VTSNDISELT----- 1579  
 Db 1733 P-----QVSAITDVEGLSDALRTEENNEGTRVTTEEPADMPASVSGDDSLTASRSEK 1788  
 QY 1580 -----STLGKHFK---DSATT-----KMLAALKE-----LDDAK 1605  
 Db 1789 DECAMISTSIGEEFELPISSATTIKCAESLOPVAVAVEERATGPVLISTADFGPMPSAP 1848  
 QY 1606 PAEQLHILQOHSKDVGDERYEAVRNKLVIROAADSHMELGSGASHSTYNNLSR 1665  
 Db 1849 PAESPLASTSKEEKDECA-----LISTSIAECEASVSGVVSENERAGTVMEE 1899  
 QY 1666 INNDGIVELLKHFDAAALPASSAKRLGEMMNDPALKDIIKQLOSTPF-----SSASVSM 1721  
 Db 1900 KDGSGLISTSSVE-DEGVPSSAVPOEE---GDPSV-----TPAEMGDTAMISTS 1946  
 QY 1722 LKDLREOTEKAILDKGVGREEVGLFDORNNL---RVKSVSVSQSVKSEGTTPALL 1778  
 Db 1947 TSEGC-----EAVM-----IGAVLODEDRLTITRVEDLS-----DAI 1980  
 QY 1779 GTSNSAAMSMERNICTINFKYQODQNTPRFTLEGGIAQANQVASAL---TDLKKEGLEM 1836  
 Db 1981 STSTAECMPISASI-----DRHEENQLTAD-----NPEGNGDLSATEVSKHKVPM 2025  
 QY 1837 KS 1838  
 Db 2026 PS 2027

## RESULT 8

PCT-US01-03782A-284  
 ; Sequence 284, Application PC/TUS0103782A  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hyseq, Inc.  
 ; TITLE OF INVENTION: Novel Bone Marrow Nucleic Acids and Polypeptides  
 ; FILE REFERENCE: 21272-040  
 ; CURRENT APPLICATION NUMBER: PCT/US01/03782A  
 ; PRIOR FILING DATE: 2001-02-05  
 ; PRIOR APPLICATION NUMBER: 09/496,914  
 ; PRIOR FILING DATE: 2000-02-03  
 ; PRIOR APPLICATION NUMBER: 09/598,075  
 ; PRIOR FILING DATE: 2000-06-20  
 ; PRIOR APPLICATION NUMBER: 09/620,325  
 ; PRIOR FILING DATE: 2000-07-19  
 ; PRIOR APPLICATION NUMBER: 60/250,583  
 ; PRIOR FILING DATE: 2000-11-30  
 ; NUMBER OF SEQ ID NOS: 386  
 ; SOFTWARE: Custom  
 ; SEQ ID NO 284  
 ; LENGTH: 5447  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 PCT-US01-03782A-284

Query Match 2.0%; Score 190; DB 1; Length 5447;  
 Best Local Similarity 17.6%; Pred. No. 0.058;  
 Matches 415; Conservative 311; Mismatches 806; Indels 820; Gaps 111;  
 QY 27 ALQOQSS-----SSSPQNAASLAA-----EG---KNRGKM-PRTHQPSTAADGISAHQ 72  
 Db 1601 ALQKNSDLKLDQDDIQNRATSFATVVKVDIEGPMEEQTKLSR--ELTALREKLQAKE 1658

QY 73 OKKSFSLRGLCTKKFSRPAQO-QPGTTHSKGATLRLDILLARDGGTQHEAAAPDAARLT 131  
 Db 1659 QYEAQOETRAQKELEAEAVTSALQOETEKSAK--KELA---ENKKKIDALLDWMVTSVG 1713  
 QY 132 RSGGVKRRN---MDDMAGRPVMKGG-----SGEDKVPTQ-----OKRH-----QL 168  
 Db 1714 SSGGOLLTNLPGMEOLSGASLEKALDITDGYMGVNOAPEKLDKQCEMMKARHOELLSSQ 1773  
 QY 169 NNF---GQMROTMLSKMAHPASANAGDRLOHSPPHIPGSHHEIKEEPVGSSTKA-TTAHA 224  
 Db 1774 QNFILATSAQAFLDQHGHNLTPEEQMLQO-----KLGELKEQ--YSTSLAQSEAL 1824  
 QY 225 DRVEIAQEDDDSEFOOLHQ-----QRLARENPPQPKLVGATPISARFQPKLTAAVE 278  
 Db 1825 KOVQTLQ-DELOKFLQDHKEFESWLERSEKELENMHK---GGSSP----- 1865  
 QY 279 SVLEGTDTQSPKQOSML-----KG-----SGAGVTPLAVTLDKGLQLADPNPP 324  
 Db 1866 -----ETPLSLLKROGSFSEVDVSHKGDRLRFVTISGQKVLDMENSFKEGK-----EPS 1913  
 QY 325 ALNTLLKOTLKGDTOHYLAHASSD--GSOHLLDNKNGHLFDIKSTATSYSVLHNSHPGE 382  
 Db 1914 EIGNLVKDKLDATERYTALHSKCTRLGSHLNNLLGQYHOF--QNSADSLQAMQACEAN 1971  
 QY 383 IKGLAQAGTGSVSDGSGKISLGSQTSHNKTMLSQPGEAHRSLLTGTWQHPAGAARP 442  
 Db 1972 VEKLLSD-----TAASDPGLVLOEQLAT-----TKQ 1996  
 QY 443 QGESIRLHD---DKTHILHPELGVWQADKDTHSOLSQADGKL-----YALKDNRTL 492  
 Db 1997 LOELAEHQVPEKLOKVARDIMEGEPADPHRVQETDTSILSHFOSLSYSLAERSL 2056  
 QY 493 -----QNLDSKSEKLVDRKISYSVDORG-QVAILLDTGCRHKMSIMPSLDASPESHI 545  
 Db 2057 LOKAIAOSQSVDSLESLLQSIGEVQNEQKQVSSLSGVLOEALATNMKL----- 2108  
 QY 546 SLSLHFADAHOGLLHGKSELEA-----QSVASHGRVLVADSEKGLFSAAI 591  
 Db 2109 -----KODIARQKSSLEATREMTVRPMETADTTAAVLQGL--AEVSOREQLCL 2157  
 QY 592 PKQGDGNELMKAMPOHALDEHFGHDHQLSGFFHDDHQLNALVKNNFRQOHACPLGNDH 651  
 Db 2158 QOQEKESSLK-KULLPOAEMFEHLS-----GKLOPFMENKSRM-----LASGN 2198  
 QY 652 QFHPGNLTDALVDNOLGLHHTNPEPH-----EILDMGHLSLALQEGKLVH--- 699  
 Db 2199 Q---PDQDITHFFQOIQELNLEMEODOENLDTLEHLVTELSGCFALDLCQHQRVQNLK 2256  
 QY 700 -FDQLTKWGTGAESDCKQLAKKGLDGAAYLLK-----DGEVKRLNINOSTSIKHGT 749  
 Db 2257 DTELOKTVKEREKADSSCOEQLDEFPFKLVRTFOKWLKETGSGIPPTETSMASKELEQI 2316  
 QY 750 ENVFSL-----PHVRNKPPEPGDALQGL-----NKDKRAOAMAVI-----GVNKY 788  
 Db 2317 EHLKSLDDWASKGTLVEEINYKGTSLNLIEMITAPDSOGKTGSIILPSVSSVGSVNGY 2376  
 QY 789 LA---LTE---KGDIRSQIKPGTOOLERPAOTLSREGISGELKDTHVDHKQNLIALTH 841  
 Db 2377 HTCKDLTEIQCDMSDVNLKYEKLGVLHERQE---SLOAILNRMEEVHKEANSVLQWLES 2433  
 QY 842 ECEV-----FHOPREA-----WONGAE 858  
 Db 2434 KEEVLKSMDSMSPKTVTKAQAESNKAFLAELEONSPKIQVKYKAGLALLTYPNQSQ 2493  
 QY 859 SSSWHL-----ALPQSESKLSLDSMSHEHKPIATFFEDGSOHOLKAGGWHAYAAP 908  
 Db 2494 AENWKKIOELNSRWERATEVTVARQORQLEESASH--LACFQ-AAESOLOP--W-----LM 2544  
 QY 909 ER-----GPLAV-----GTSGSQTVF-----NRLMOGVKGKVIPGSG--- 940  
 Db 2545 EKELMMGVGLPSLDIPNMLNAOKQOVQFMKFEFARROOHEQLNEAAQG--ILTGPQDVS 2603



941 QY -----LTVKLSAOTGGMTGA-----EGRKVSCKFSERIRAYAFNPTM 977  
2604 DB STSQVQKELASINOKWVETDOKLSRSQIDQAIKSTQYQELQDLSEKVRAG-----2658  
978 QY STPRPIKNAAYATQHGQWGREGLKPLYEMOGAL-----IKOLDAHNVRHNP-----1024  
2659 DB -ORLSVQSAISTQ-----PEAVKQOLEETSEIRSDLEQLD-HEYKEAQTLCDELVSGLIG 2710  
1025 QY -----OPDLOSKEITLDLGEHGAEE-----LLNDMKRFRDELEQS 1058  
2711 DB EYKLDKELKRLTETVALPLOGLEDLAADRINRLOALASTQOQFDFELRTWLDKQSQ 2770  
1059 QY -----LGOHQG-----VLKSNGE-INSEPKPSGKALVOS 1094  
2771 DB QAKNCPIAKLERLSQLOQENEEFKSLNHSYSEYVIAEGESLILLSVPPGEERKTLON 2830  
1095 QY -----FVNRSQDLSKS-----LQQA-----VHATP 1116  
2831 DB QLVELKNHWEELSKKTADRSRLKDCMOKAQYQHVEDLVPIEDCKAKMSRLVTLDP 2890  
1117 QY PSRESKL---OSMLG-----HFVSAGVDMHQKGEIPLGRQDPNDKTAITKSLRI 1164  
2891 DB VOLESLLRSKAMLNEVEKRSLLEILNSAADILINSSEADEGIRD--EKAGINON---2945  
1165 QY LDTVT-----JGELHELADKAK-----1181  
2946 DB MDAVTEELQAKTGSLEMTQRLREFQESPKIEKKVEGAKHQLIEFDALGSQACSNKLE 3005  
1182 QY -----LVSDBKPDADQIKQLRQOQFDTLREKRYESNPVKHY 1216  
3006 DB KLRQAEVLOALEPQVDYLRFNTQGLVED-APDGSASOLLHQAEEVAQOEFL---VKOR 3061  
1217 QY TDMG---FTHNK-----ALEANYDAVKAFI 1238  
3062 DB VNSGCMVMMENLEGIGQFHCVRREMFSQLADLDLDMGAIGRDTSLOQIEDVRLFL 3121  
1239 QY N---AFKKEHGVNLTTRTVLESQSAEL-----AKKLNLTLSLD 1276  
3122 DB NTHVUKLDIEASEACRHLMEEGTLLGLKRELEALNKQCKLTERGARQOELELT 3181  
1277 QY SGE SMSFSRSYGG-----GVSTVFV-----PTLSKVPVPVPIGA 1311  
3182 DB LGRVEDFYRLKGLNDATTAAEEALQWVVGTEVEIINOOLADFKMFQKQVDEL---3237  
1312 QY GYTLDRAYNLNLSFRTSG---GLNVSFGRDGVSGNIMVATGHVPMYMTGKTSAGNASD 1368  
3238 DB ---QMKLOOVNGLGLOGLQSACKDCDVQ-----LEHDMEEI-----NAR- 3274  
1369 QY WLSAKHKIS---PDLRIGAAVSGTLOGTLQNSLKFTEDEL-----PGFIHGLTHGTL 1419  
3275 DB WNTLNKVAQRIQAQLQEAALLHCGKFQDALEPLLSWLADTEELIANOKPSPAIEYKVVKAOI 3334  
1420 QY TPABELLOKHEHQMKQSKLTFSDTSANLDRAGINLNEGSKPVGVTARVSAGLSASA 1479  
3335 DB QEQKLLQRLD-----DRKATVDM-----LQAEGR-----IAQSA 3365  
1480 QY NLAAGSRSRSTSGOFSSTTSASNNRPTFLNGVGGAGANLTAALGVYAHSTHIE-GRPVGIF 1538  
3366 DB ELA--DREKIT---GOLESIESRWTELLSKAAARQOELELVLAKOFHETABPISDF 3418  
1539 QY PAFSTINVSAAALADNRFTS---QSTISLELKRAEPTVSN-----DISELSTSLIG 1583  
3419 DB LSVTEKKLANSEPVGTOTAKIQOQIIRHKALFEEDTENHATDVHQAVKVGQSLSLTSAPAE 3478  
1584 QY KFKDSATTKMAALKELDDAKPABOLHLOQHFSAKDVVGDERYEA-----1630  
3479 DB QGVLESEKIDSQARYSEIQD-CCCKKAALLDQALSNAFLFGEDEVILNWLAEVEDKLSS 3537  
1631 QY ---VRMLKLVIRQQAAD-----SHSMELGSA--SHSTTYNNLSRINN--DG 1670  
3538 DB VEVKDFKQDVLHROHADHLALNEEIVNRKNKYDQAIKNGOALLKQOTGEVLLIQEKLQD 3597  
1671 QY IVELLHKHFDAAIPASSAKRLGEMMNDPALKDIILKQLOSTPFPSSASVSMELKGLREQT 1730

3598 DB IKT---RYADITVTSKALRTLEQ-----AROLATKFOST-----YELTGLWREVE 3641  
1731 QY EKAILDKGVGREVG---VLFDQRNNLRKVSQSVSKSEGENTPALLLGTNSAAMS 1787  
3642 DB EELATSG---QOSPTGQIPOFQORQKELKKEVMEHRLVLDTVNEVSRALELVPRAREG 3699  
1788 QY MERNIGTINPKY 1799  
3700 DB LDKVSDANEQY 3711

RESULT 9  
US-09-196-296B-2  
; Sequence 2 Application US/09196296B  
; GENERAL INFORMATION:  
; APPLICANT: Suen, Chen-Shian  
; APPLICANT: Fraill, Donald E.  
; APPLICANT: Lyttle, Richard C.  
; TITLE OF INVENTION: Cloning and Expression of a Nuclear  
; FILE OF INVENTION: Receptor Coactivator Proteins and Uses Thereof  
; FILE REFERENCE: 0630/01376  
; CURRENT APPLICATION NUMBER: US/09/196,296B  
; CURRENT FILING DATE: 1998-11-19  
; NUMBER OF SEQ ID NOS: 3  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 1423  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-09-196-296B-2

Query Match 2.0%; Score 188.5; DB 5; Length 1423;  
Best Local Similarity 17.8%; Pred. NO. 0.0064;  
Matches 239; Conservative 190; Mismatches 421; Indels 493; Gaps 62;

681 QY ILDMGHLSLALQECGLHYFDQLTGWTGAESDCKQLKKGLDGAAYLL-KDGEVKRLNIN 739  
107 DB VIDKSLGLPLLQ-----ALDGLFVNRDGNIVFVSEN 140  
740 QY OSTSIKIHGTE-----NVFSLPH-----VRNKEP-----763  
141 DB -VTQYLOYKOEDLVNTSVYNILHEEDRKDFLKNLPKSTVNGVSWTNETQKSHTFNCRM 199  
764 QY -----PGDALOGLNKD---DKAQAAMAVIGVKNYLAITEKG-DIRSF-----QIKPGTQQ 808  
200 DB LMKTPHDIILEDINASPENRQRYETMOCFALSQPRAMMEGEDLQSCMICVARITTCERT 259  
809 QY L-ERPAOTLSREGISGELKDIDVHK-----ONLYALTHEGEVFPQPR- 850  
260 DB FFSNPESFITRHDLSGKVVNIDTNSLRSSMRPGFEDIIRRCIORFFSL-NDGQSWSQKRH 318  
851 QY ---FAMONGAESSSWHKAL-----PQSEKLSLSDMSHEHKPIATFEDG--SOH--OL 897  
319 DB YQEAYLNGHAETPVYRFSLADGTIVTAOTKSKL-----FRNPTVNDRHGFSVTHFLOR 371  
898 QY KAGGHAYAAAPERGLAVGTSGSOTVFENLMQGVKGVIPSGSLTVKLS-----AQO 949  
372 DB EQNGYR-----PNPNVCOGIRPPMAGCNSSVGGM--SMSPNQGLOMPSSRAYGLADPSTT 425  
950 QY GGMTGAE--GRKVSCKFSERIRAYAFNPTMSTPRPIKNAAYATQHG--OQREGKLPYE 1005  
426 DB QWMSGARYGG-----SSNIASLTTPGQMOSPSYQNNNYGLNMSPPHSGSLGLAP--N 476  
1006 QY MQGALIKQLDAHNVRHNAPODLSQKLETLDLGEHGAELLNDMKRFRDELEQSATRSVTV 1065  
477 DB QQNIMTSP-----RNRGSPKIAH-----QFSPV 500  
1066 QY LGHQHGVKLSNGEINSEFKPSGKALVQSFVNRSGQDLSKLSLOQAQVHATPPSAESKLOS 1125  
501 DB AGVHSPMASSGNTGNHFSSSLSAL-----QAISEGVGTSLTSLTSSPGPKLDN 550





Db 2268 KSL--LDMASKGTLVEIN--CKGTSLE---NLIMEITAPDSOGKTDLTETIOCDMSDV 2319  
 QY 808 QL--ERPAQTL--SREGISGELKDTHVHKQNLIALTHEGEV-----845  
 Db 2320 NLKYEKLGVLHERQESLQALINRMEEVHKANSVLOWLESKEEVKMSDMASSTPKTET 2379  
 QY 846 -----FHOPREA-----WONGAESSWHKL-----AL 867  
 Db 2380 VKAQAESKAFIAELQNSPKIQKVEALAGLLVYPNSQEAENKKIQEELNSRWERAT 2439  
 QY 868 POSSEKSLASLMSHSHKPIATFEDSGHQLKAGGHAYAAPR-----GLAV-----915  
 Db 2440 EYTVARQLESASH--LACFO-AAESOLRP--W--LMEKELMMGVGLPUSIDPNML 2490  
 QY 916 GTSGSQTVP-----NRLMQGVKGKVLPGSG-----L 941  
 Db 2491 NAAQOQVFMLEKFEARRQOHEOLNAAQG--ITPGDVSLSTSQVQKELQSNQKVEL 2549  
 QY 942 TVKLQAQTGGMGA-----EGRKVSFKSERIRAYAFNPTMTSPRIKNAAYATQHWQW 996  
 Db 2550 TDKLSRSSQIDOAIVKSTQYQELLQDLSEKVRAGV-----QRLSVQSAISTQ-----2597  
 QY 997 REGPLXYEQGAL-----IKOLDHNVHNP-----OPDLQSKLETLDLGE 1039  
 Db 2598 PEAVKQOLETESEIRSDLEQLD--HEVKEAQTLCDELSVLGEQYLKDLKKRLTVALPL 2656  
 QY 1040 HGAE-----LLNDMKFRDELEQSATRSVTV-----1065  
 Db 2657 QGLDLAADRLNLAALASTQFOQMFDELRTWLDKQQAQKNCIPISAKLERLSOLQ 2716  
 QY 1066 -----LCQHOG---VLKNSGE-INSEFKPSGKALVOS--FNVNRSGQDLSKS-----1107  
 Db 2717 ENEEFQKSLNQHSGSVIVAEGESLLSVPPGEEKRTLQNLQVLEKLNHWEELSCKTADR 2776  
 QY 1108 -----LQQA-----VHATPPSAESKL---QSMGL-----1128  
 Db 2777 QSLKDCMQAKQYOHVEDLPVWIEDCKAKMSELRTVLDPVQLESSLLRSKAMLENEVK 2836  
 QY 1129 -----HVSAGVMSHKGEPILGRQDPNDKTAITSRLTDTVT-----IGELHELA 1177  
 Db 2837 RRSLETLNSAADILINSSEADEGIRD--EKAGINQN--MDATVEELQAKTGSLEEMT 2891  
 QY 1178 DKAK-----1181  
 Db 2892 QRLREFOESFKNIEKKVEGAKHOLEIFDALGSAQCSNKNLEKLRQAEVLOALEPOVDYL 2951  
 QY 1182 -----LVSDHKPADQIKLRQOQDFTLREKRYESNPVKHYTDMG--FTHNK-----1225  
 Db 2952 RNFTQGLVED--APDGSASQLLHQAEVACQEPLE---VKQVNSGCVMMENKLEGIGQPH 3007  
 QY 1226 -----ALEANYDAVKAFIN---AFKKEHGVNLTFT 1254  
 Db 3008 CRVREMFQSLADLDELDMGMAIGRDTDSLSQIIEVRLFLNKHVILKLDIEASEACRH 3067  
 QY 1255 VLESQGSAL-----AKKLNTLLSIDSSESMSFSRSYGG-----1289  
 Db 3068 MLEEGTLDLGLKREALEALNKCQKLTGERKARQOQLETLGRVEDFYKRLGLNDATT 3127  
 QY 1290 -----GVSTVFY-----PTLSKKVPVPVIFCAGITLDRAYNLSFRTSG--1328  
 Db 3128 AAEAEALOWVYGVTEIINOQLADKFMKQKEQVDP-----OMKLOQVNLGL 3175  
 QY 1329 -GLNVSFRGDSVGNIMVATHDVMPYMTGKTSAGNASDWLSAKHKIS---PDLRIGA 1384  
 Db 3176 QGLISAGKDCDQV-----LEHDMEEI-----NAR-WNTLNKKVQAORTAQLOEAL 3220  
 QY 1385 AVSGITQTLQNSLKFLELDEL-----PGFTIHLTHGTLPALQLQKGIHOMQOQSK 1438  
 Db 3221 LHCGRFQDALEPLLSLWLADETEELIANQKPPSAEYKVKVQAQIQEQLKLRLLD-----3272  
 QY 1439 LTFVSDTISANLIDLRAGINLNEGSKPNVGTARVSAGLSASANLAAAGSRSTTSGFGST 1498  
 Db 3273 -----DRKATVDM-----LQAEGR-----IAQSAELA--DREKIT-----GOL 3304

QY 1499 TSASNNRPTFLNGVGACANLTAALGVAHSSTHE-GKPVGIPPAFTSTNVSAALADNRST 1557  
 Db 3305 ESLESRTWELTSKAAAKOKOLELILVLAQFHETABEIPISDFSVTEKKLANSEPVGTOTA 3364  
 QY 1558 --QSISELKRAEFTSN-----DISELSTLCKHKFKDSATTMKLAALKELD 1602  
 Db 3365 KIOQIIRHKALEDIENHATDVHQAQVIGOSLSLSLTPAEQVLSERKIDSLQARYSEIQ 3424  
 QY 1603 DAKPAEQHLHLOQHFSAKDVVGDEREYEA-----VRNKKLVIRQQAAD--1645  
 Db 3425 D-RCRKAALLDQALSNAFLGEGEDEVENLWLAEVEDKLSVVFVKDFQDVLHRQHADHL 3483  
 QY 1646 -----SHSMELGSA-SHSTTYNNLSRINN--DGIVELLHKKHFDALPASSAK 1689  
 Db 3484 ALNEEIVNRKNVDQAINKGOALLKQTGTGEEVLLIQEKLDTKT---RYADTIVTSSKAL 3540  
 QY 1690 RLGEMNNDPALKDIIKOLQSTPSSASVSMELKDLGREQTEKAILDGVGREEVG---V 1746  
 Db 3541 RTLEQ-----ARQATKFOST-----YEELTGWLREVEELATSG--GQSPTGEQIP 3585  
 QY 1747 LFQDRNNLRVKSVSQSVSKSEGFENTPALLLGTNSAAMSERNITGTFNKY 1799  
 Db 3586 QFOQRQKELKKEVMEHRLVLDIVNEVSRALLELVPRAREGLDKLVSDANEQY 3638

RESULT 11  
 US-08-945-567C-4  
 ; Sequence 4, Application US/08945567C  
 ; GENERAL INFORMATION:  
 ; APPLICANT: SASAKI, Ken  
 ; APPLICANT: HARKNESS, Robin E.  
 ; APPLICANT: LOOSMORE, Sheena M.  
 ; APPLICANT: CHONG, Pele  
 ; APPLICANT: KLEIN, Michel H.  
 ; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE PROTEIN OF  
 ; FILE REFERENCE: 1038-745 MIS  
 ; CURRENT APPLICATION NUMBER: US/08/945,567C  
 ; CURRENT FILING DATE: 1998-03-19  
 ; PRIOR APPLICATION NUMBER: 08/431,718  
 ; PRIOR FILING DATE: 1995-05-01  
 ; PRIOR APPLICATION NUMBER: 08/478,370  
 ; PRIOR FILING DATE: 1995-06-07  
 ; PRIOR APPLICATION NUMBER: 08/621,944  
 ; PRIOR FILING DATE: 1996-03-26  
 ; PRIOR APPLICATION NUMBER: PCT/CA96/00264  
 ; NUMBER OF SEQ ID NOS: 10  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 4  
 ; LENGTH: 1833  
 ; TYPE: PRT  
 ; ORGANISM: Moraxella catarrhalis  
 ; US-08-945-567C-4

Query Match 2.0%; Score 184.5; DB 4; Length 1833;  
 Best Local Similarity 18.3%; Pred. No. 0.018;  
 Matches 284; Conservative 214; Mismatches 537; Indels 521; Gaps 71;

QY 539 ASPESHISLSLHFADAHOG---LLHKSELEAQSVAISHGRVLVADVSEK--LFSAAIPK 593  
 Db 16 ATAKSAVSLVGLATAAGTQSTIAIGSDATSSSLGAIALGAGTQALOGSIALGQGSVVT 75  
 QY 594 QGDGNELKWKAMPQHALDEHFHGHDOISGFFHDDHGNALVKNVNFROQHACPLGNDHQF 653  
 Db 76 QSDNNSRPAYPTNQALDPKFOATNTKA-----GPLS--IGSNSIKRKIIING-----122  
 QY 654 HPGWNLTALVT-----DNOLGLHTNPPEHIELDMGLHS 689  
 Db 123 -AGVNTKDAVNVAQLEAVVVKAKERRITFGDDNDNDTDVKGILDNT-----166

```
QY 690 LALQEGKLYFDOLTKGTGTA--ESDCKQLKGLDGAAYLLKDGVEVKRLININOSTSSIKH 747
Db 167 LTIKGA--ETNALDNNIGVVKEADNSGLKVKL--AKTLNNLTETVNTTLN--ATTIVKV 221
QY 748 G-----TENFSLPHVRNKPEFGD-----ALQGLNKDDKAQAMAVIGVKNYLALTEK 794
Db 222 GSSSTTAELLSLSITFPQNTGSGQSTSKTVYGVNGVKFTNNAETAAIGTR-----274
QY 795 GDIRSEIOPCTOQLERPAQTLREGISGELKDIHVDHKQNLIALTHEGEVHFOPREAWO 854
Db 275 -----ITDKI--GFARDGDVDEKQAPY-----295
QY 855 NGAESSWHKALPQSESKLSLSDMSHEHKPIATFEDGSOHLKAGWHAYAAPRGPIA 914
Db 296 -----LQKOLKVGSAI-----TIDNG-----IDAGN-----KKISNLA 325
QY 915 VGTSGSQTVPNRLMOGVKGVIPQSGLTU-----KLSAQTGGMF-----CAE 956
Db 326 KGSANDAVTIEQLKAAKPTLNAGAGISVTPTEISVDAKSGNVTAPTYNIGVKITELNSD 385
QY 957 GRKVSFSEIRIRAYAFNPTMSTPRPI-----983
Db 386 G--TSDFK--VKGSGTNNSLVTAHSLASVNLNVRNRTADSALQSFTVKEEDDDANAITV 441
QY 984 -----KNAAYATOHGQGRGLKPLYEMOGALIKOL-----DAHNVRHNA 1023
Db 442 AKDITKNAGAVSILKLGKNGLTATVATKDGTVTFGLSQDSGLTIGKSTLNDGLTVKDTN 501
QY 1024 POPDLOS---KLETLDELGEALLNDMKFRDELQOSATRSVTVLGOHGVKLSNGEIN 1080
Db 502 EQIOVCANGIKFTNVNGSPGTGIANTRITRDKIGFAGS-----DGAVDNKNPYL 552
QY 1081 SEFKSPGKALVOSFNVRNSGODLSKSLQAQVHATPPSAESKLQSLMGLHPVSAGVDMSHQ 1140
Db 553 DQDKLQGVNKTINTGINAGGRAIT-----GLSPTLPSI-----AQSSR 592
QY 1141 KGEIPLGRORDNDKALTAKSLILDT-----VTIGEL 1173
Db 593 --NIELONTIQDKKSNASAINDIILNTGFLNKNNNPIDFVSTYDIDVFANGNATTATVT 650
QY 1174 HELADK--AKLVSD-----HKPDADQIKOLRQOFDTLREKRYENPVKHYTDMGFTHN 1224
Db 651 HDTANKTSKVYDVVNDVDTTILHTGDDNKKLGVTTLKNTKSANGN-----TATFNFN 705
QY 1225 KALE---ANYDAVKAFINAFKEHGVNLTTRTVLESQGSABELAKKLNTLLSDSGEM 1281
Db 706 SDEDAVNADIAENLATLAKETHTTKGTADTALQT-----FTVKKVDENNADANAIT 761
QY 1282 SFSRSGGVSTVFPTLSKVPVPIPCAGITLDRAYNLSPS--RTSGGLNV--SFGRDG 1338
Db 762 VGOKNANNQVNTL---TLKGE-----NGLNKTDKNGTVTFGINTTGLKAGKSTLNDG 812
QY 1339 GVS-----GNLWATGHVDPYMTGKTSAGNASDWLSAKHKISPDLRIGAAVSGTLOQT 1393
Db 813 GLSKNTPTSGQIOVGADGVKFA-----KVNNGVVGAGIDGTT---851
QY 1394 LONSLKFKLTEDELPGFTGHGTLHTLPAE--LLQKGIEHQMKQSKLTFVSDTTSANLDL 1451
Db 852 -----RITRDEI-GFTG--TNGSLDKSPHLSKDGIN---AGCKKIT-----NI 889
QY 1452 RAG-INLINE-----DGSPKNGVTARVSAGLSANLAAGSRSTRTSGQFGSTTSASNNRP 1506
Db 890 OSGEIAQNSHDVATGGKIDYLDKLELNKISSAKTAKNSLHEFSVADQGNFTVSNPYS 949
QY 1507 TF-----LNGVVGAGANLTAALGVASHSTHEKRPVGIFFAPFTSTNVSAALAD 1553
Db 950 SYDTSKTSOITFAGENGITTKVN-----KGVVRVIGDOTKGLTTPKLTGVNNGKGVIVID 1005
QY 1554 NRTSQSISLELKAEPVTSNDISELTSLGKHFEDSA---TTKMLAALKELDDAKPAEQL 1610
Db 1006 SONGQ-----NTITGLSNTLANVTNDKGSVRTTEOGNIKEDDKTRAASIV 1051
QY 1611 HILQOHFS-----AKDVVDERYEAVRNJLKLVIROOAAADSHSMELGSAHSTTYNNLS 1664
```

Db 1052 DVLSAGENLOGNGEAVDFV---STYDTV-----NFADGN-----ATTAKVTYDDTS 1094

QY 1665 RIN-----NDGIVELLHKHF-----DAALPAS-----1686

Db 1095 KTSKVYDVVNDVDTTIEVKDKKLGKVTTLTSTGTGANKFPALSNOATGDALVKASDIVAH 1154

QY 1687 -----SAKRLGEMNN-----DPALKDIIKLOSTPFSASVSMELKDGRLRQTEKA 1733

Db 1155 LNTLSGDIQTAKGASQANSAGYVDADGNKVIYDSTONKYQAK-----NDGTVDKTRV 1209

QY 1734 ILDGKVGREEV--GVLFQDRNNLRVKSVSOSV---SKSEGFNTP--ALLGTSNSAAMS 1787

Db 1210 AKDKLVAQAQTPDGTLAQ-----MNVKSVINKEQVNDANKKOGINEDNAFYKGLEKAAASDN 1265

QY 1788 MERN-----IGTIN-----FKYGODONTNPRR-----FTLEGGIAQAANPOVASALTD 1828

Db 1266 KTKNAAVTVGDLNNAVAQAQTPLTFAQDGTGTTAKKLGETLITKGGQDTN-----KLTD 1316

RESULT 12

PCT-US01-04098A-1679

Sequence 1679, Application PC/TUS0104098A

GENERAL INFORMATION:

APPLICANT: Hyseq, Inc.

TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides

FILE REFERENCE: 21272-029

CURRENT APPLICATION NUMBER: PCT/US01/04098A

CURRENT FILING DATE: 2001-02-05

PRIOR APPLICATION NUMBER: Not Yet Assigned

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: 09/728,422

PRIOR FILING DATE: 2000-11-30

PRIOR APPLICATION NUMBER: 09/693,325

PRIOR FILING DATE: 2000-10-20

PRIOR APPLICATION NUMBER: 09/663,561

PRIOR FILING DATE: 2000-09-15

PRIOR APPLICATION NUMBER: 09/654,936

PRIOR FILING DATE: 2000-09-01

PRIOR APPLICATION NUMBER: 09/620,325

PRIOR FILING DATE: 2000-07-19

PRIOR APPLICATION NUMBER: 09/598,075

PRIOR FILING DATE: 2000-06-20

PRIOR APPLICATION NUMBER: 09/560,875

PRIOR FILING DATE: 2000-04-27

PRIOR APPLICATION NUMBER: 09/496,914

PRIOR FILING DATE: 2000-02-03

NUMBER OF SEQ ID NOS: 3960

SOFTWARE: Custom

SEQ ID NO 1679

LENGTH: 2189

TYPE: PRT

ORGANISM: Homo sapiens

PCT-US01-04098A-1679

Query Match 1.8%; Score 167; DB 1; Length 2189;

Best Local Similarity 18.8%; Pred. No. 0.32;

Matches 373; Conservative 224; Mismatches 754; Indels 634; Gaps 91;

QY 89 SRAPGQPGT-----THSKGATLRDLLARDGGETQHEAAAPDAARLTRSGGVKRRNMD 142

Db 412 SAEAPKSGPSGVMKKELSGSSSAPKLEYTVRTDTQSTPTNGSPSSPO--OKSEGLGSRHRP 470

QY 143 DNAGRPWKGGSGEDKVPQQ-----KRHQLNNE-----GOMROT 177

Db 471 VARVSPHCKRSEAEAKPSGSGTQVNLTRGRANDPCDLDSRVQATSVKVTAGFQPGGAVEKE 530

QY 178 MLSKM-----AHPASANAGDRLOQSPPHIPGSHHEIKPEP-----VGSTS 217

Db 531 SLGKLTTCAGCVSTSCELASALSHLDASHLTENLPRAASELGOOPMTFELDSSSLSSPG 590

QY 218 KATTAHADRVETAQEDDSEFOOLHOORLARERENPPQP--PKLG-----VATPISAR 268

Db 591 KGAHPDPKTSVDTG-----KVRP-ENPSQAPRVAKCAKSPVRLPHEGS 639  
 Qy 269 FQPKLTAFAE---SYLEGTDTQSP--LKQSMKLGSG---AGVTPLATLDKGLQOLAP 320  
 Db 640 PSPGEKAAAPDYSKTRGASSETSTPHNTRRVAALRGAGGEMTPAGAVL-----P 691  
 Qy 321 DNPALNTLLKQTLKQTOHYLA-----HHASSDGSQHLNLDNKGHLFIKSTATSYSL 375  
 Db 692 GDP-----LTSQORQAPGNHSALEMTGTHAPESQPSLLEGADSV--SSRAPOASLSML 747  
 Qy 376 HNSH-----PGE-----IKGLAQAGTGSVSDGSKG-----403  
 Db 748 PSTDNTKEACGHVSGHCCPGGSRSPVTDIDSFIKELDASAARPSQTDGSGSQSAQ 807  
 Qy 404 ----ISLGGTOSHNTML--SQGEAHRSLLTGIMOHAPAGAARPOGESIRLHDDKIHL 458  
 Db 808 GHPPAGAGGSSCRAPVPGGOTSSPRRAWAGAPAYPOWASOP-----851  
 Qy 459 PELGYWOSADKOTH-----SOLSRQADGKLYALKDNRTLQNLNDKNSSEKLVKIKSYV 513  
 Db 852 ---SVLDSINPKHFTVKNFNLSNYRNFSSPHEDSTSLGLGD---STEPSSLSSMYGDAE 906  
 Qy 514 DORGQVAILTDP--GRHKMSIMPSLDASPEHSISLHFADAHGOLLHGKSELEAQSV 570  
 Db 907 DSSDPESLTPAPRASARDGWS-----PPRSRV--SLHKED-----PSESEEEQI 949  
 Qy 571 AISHGR-----LVVADSEGLF-----SAAIPKO-----GDGNEI 600  
 Db 950 EICSTRGCPNPPSPAHLPQAAICPASAKVLSLKYTTPRESVASPREKAACLPGSTSG 1009  
 Qy 601 KMKAMPOHALDEHGHQHSIGFHHQGL-----NALV 635  
 Db 1010 PDSSQPSLL-----EMSSQEHETHADISTSONHRPSCAEETETVSASSAMENSPL 1061  
 Qy 636 KNFROOHACPL-----GNDH-----QHPGWNLTAL-----V 664  
 Db 1062 SKVARHFSPPIILSSPNMNGLEHLLDDETLNOYETSINAAASLSFSVDVPKNGESV 1121  
 Qy 665 IDNOLGHTNPEPEHILDMHGLSLALOEKLYHFDQLTGWTGAESDCKOLKGLDGA 724  
 Db 1122 LEN---LHIS--ESQDLDDLQPKMTARRPIMAFKEINKHNOGTHLRSRTEKE-----1171  
 Qy 725 AYLLKDEVKRLNINQSTSSIKHGTENVFSLPHVRNKPDPGALOG---LNKDDKAQAMAV 782  
 Db 1172 -QPLMPARSPDKTOMVSSSKGV---TVPH--SPQPKTNLENKDLSSKSPAEMLLT 1224  
 Qy 783 IGVN-----KYLALTEKGDIRSFOIKPGTQOOLRPPACTLSREGISGELKDI---HVHD 832  
 Db 1225 NGQAKCGPKLRLSLKGAKVNS-----EAPAAVAKAGGTDHHRKPLISQPTSH 1274  
 Qy 833 K-----ONLYALTHEGEVHPQPREANQNGAESSWHKLALPOS-----ESKLKSLDMS 880  
 Db 1275 KTLKAVSQRLHVADH-----EDPDRTNTAAPRSPQCVLESK--PPLATS 1317  
 Qy 881 HEHKP-----IATED---GSQHQLKAGGHAY-----AAPERGLAVGTSGSQTVFNR 926  
 Db 1318 GPLKFSVSDTSRTFVSLTSPKVPQEQMWSRFHMAVLSEDPDRCGCTTPKSPKCAEGR 1377  
 Qy 927 LMQGVKGVIPGSGGLTVKLSAQTGGMWGAEGRKVSXKFSERIRAYAFNPTMTSPRI--K 984  
 Db 1378 APRADSGPVP-----AASRNGMSVAGNKOSEPLASHVAA-----DTAQPRTEGEK 1424  
 Qy 985 NAAATQHCWOCREGLKPLYEMOGALIKOLDAHNVHRNAP-----QPDLSKLE 1033  
 Db 1425 GGNIMASDRLENTNLK-IVEISAEVSE-----TVCNKPFAESDRRGCLAOGNCOEKE 1479  
 Qy 1034 T---LDLGHGAELNDMKRRFRDELEQSATR--SVTVLGOHQGVLYKNSGEINSEFKPSG 1088  
 Db 1480 IRLYQVAESSTSHPSLPSHSAQAEQMSRFSMTKLASSSSSLQT-ATRKAEYSOGKS 1538  
 Qy 1089 KALVQSFNNRSQDLKSLQOQAVHATPPSAESKLOSMGLGHFVSAGYDMSH--QKGEIP- 1145

Db 1539 SLMSDSRGVPRNSIPGGPSCGEDHLYFTPRPA-TRTYSMPAQF-----SSHFGREGHPH 1591  
 Qy 1146 -LGRQDPNDKTKALTSLRLDITVITIGELHELA-----DKAKLV 1183  
 Db 1592 SLGRSRD--SQVPVTSVVPPEAKASRGGLPSLANGQGIYSVKPLDLDTSRNLPADEGDII 1649  
 Qy 1184 SDHKPD--ADOIKQLRQOQDITLREKRYESNPVKHYTDMGFTTHNKALANDYDAVKAFINA 1240  
 Db 1650 SVQETSCLVTDKIKVTRHY-----CYEQN-----WPHE---STSFESVKORIKS 1691  
 Qy 1241 FKXEHGVNLT--RTVLESQGSAEALAKKLNTLLSLDGSMSFSRSVGGGVST-----1293  
 Db 1692 FE-----NLANADRPVAKSGASPFLSVSSKPPGIRRSSGSIYSGSLGHPGDAARLLRR 1745  
 Qy 1294 -----VFVPTLSKKVPVPIPGAGITILDRAYNLSFRTSGGLNYSFGRDGV 1340  
 Db 1746 SLSCSENQSFAGILLPOMAKS---PSIWTLTISRQNPETSSKGSSELSKSLGLPLGP 1802  
 Qy 1341 SGNIWATGHDVMPYMTGKKT-----SAGNAS 1367  
 Db 1803 TPTMTLAS-----PVKRNKSSVRHTQPSVSRSKLOELRALMPDLKLCSEDSYSGPSA 1857  
 Qy 1368 DWLSAKHKISPDRLRTGAASVGTLOGTLONSLAKFKLTDELPGFIHGLTHGLTLPALLQK 1427  
 Db 1858 VLFKTELEITPRSPGPPAGGV-----SCPEKGNRACPGSGGPKTSAETPSSASDT 1910  
 Qy 1428 GIEHOMKQSKLTFVSVDTSANLD--LRAGIN-----LNEGSKPNGVT-----1468  
 Db 1911 G-----EAAQDLFFRRSVSNVLDLIVSAGDOORLOSVSSVSGSKTILTLIOEAKQSE 1965  
 Qy 1469 -----ARVSAGLSASANAAGSRSTTSQFGSTTSASN-----NRPTFLNGVG 1513  
 Db 1966 NEEDVCFIVLNKESGLGFSVAGGTDVPEKSIIVHVFSGGAASQEGTNNRGDFLLSVN 2025  
 Qy 1514 AGANLTAALGVVAHSST-----HEGKPVGIFPAFTSNVSAALADNRTSOSISLEL 1564  
 Db 2026 -GASLA--GLAHGNVLKVLHQAHLKDALVWIKGMDQPRPSA-----2065  
 Qy 1565 KRAPVTSN-----DISELTSTLGHKF---KDSATTK 1593  
 Db 2066 -RQEPPTANGKGLLSKRTIPLEGIGRSVAVHDAALCEVLTKSAGLSLDGGKSSVTGD 2124  
 Qy 1594 MLAALKELDDAKPAEQHLILQ-----HFSKDVGVDEREAVRNK-----KLVIRO 1641  
 Db 2125 GPLVIRKVGKGVAEQAGIIEAGDEILAINKPLVGLMHFDWAIMKSVPEGPVOLLIRK 2184  
 Qy 1642 QAADS 1646  
 Db 2185 HRNSS 2189

RESULT 13

US-09-739-449-8806

Sequence 8806, Application US/09739449

GENERAL INFORMATION:

APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.

TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof

FILE REFERENCE: 38-10(15490)C

CURRENT APPLICATION NUMBER: US/09/739,449

CURRENT FILING DATE: 2000-12-19

PRIOR APPLICATION NUMBER: US 09/514,000

PRIOR FILING DATE: 2000-02-23

NUMBER OF SEQ ID NOS: 13351

SEQ ID NO 8806

LENGTH: 1228

TYPE: PR

ORGANISM: Agrobacterium tumefaciens

US-09-739-449-8806

Query Match

Best Local Similarity 1.7%; Score 164; DB 5; Length 1228;

18.6%; Pred. No. 0.17;

Matches	282;	Conservative	185;	Mismatches	513;	Indels	536;	Gaps	65;
---------	------	--------------	------	------------	------	--------	------	------	-----

Qy	488	DNRTIONTSDNKSEKLVDIKSYSVDORCOVAILLTDTPCRHKMSTMPSLDASPESHISL	547
Db	28	DRSTLWNAE-----FAEKRRDARVAREFETAL-----PHLESPGELKAARAFQODI	75
Qy	548	SLHFADAHOGLLHGKSELQAOSVALISHGLRVYADSEGKFLSAAIPKQDGNELKKMKAMPO	607
Db	76	NNRYCAAADFALHSPSE-----HGDIRNHHAHVLMTTPOVGKAGLCERTCL-----	121
Qy	608	HALDEHFHDHOISGFDDHGHQLNAL-----VKNNFRQQHACPLGNDOHPFGCNLTLD	661
Db	122	----EHKNARILANGMATMD-QLRDIRQSWEGINARLOLHEGLDVRIDHRSH-----	169
Qy	662	ALVIDNOLGLHTNPPEHILDMCHLG-----SLAQBGKLUHYFDQUTKGWTGAES	712
Db	170	--IERGLELSPT-----HMGVHASOMROQGMAVERGL-----	201
Qy	713	CKOLKGLDGAAYLLKDGEVKRLNINOSTSIKHGTENVFSL-PHVRNKPPEGDALOGL	771
Db	202	-----DDEAAR-----QNAALIRQKPOVLTLISHEKSVDFRDHIAKTL	240
Qy	772	NK--DDRQAQ-----AMAVIGNKYIALTEKGDIRSFQIKPGTOQLERPAQTLSREGISGEL	825
Db	241	HRYINDAOQTFONAFAAV-----LASSALVELQAEIRDPCGTGSVN-ARYSTREMIDLEL	294
Qy	826	KDIVDHKQNLYALTTHECEVFHQPREAWQNGAESWHKLALPOSES-KLKSLDMSIEHK	884
Db	295	-----AMARSAVRLHQ--AOSHGVDPRHVDRAIERODRSLRSGSGCMLAASD	339
Qy	885	PIATFEDGSOHLKAGWHAYAAPERPGLAVGTSGS--QTVPN--RLMQGVKGVKVIKPSG	940
Db	340	PSAGLSDOEORHAIK-----HITGSRIVAVVGVFAGAGKSTMLTAAARKAWEAQGVYHGAA	394
Qy	941	LTVKLSAQTGMGTGAEGKRVSSFSERI RAYANPMTSTPRPKNAAYATOHW-QRGRE	999
Db	395	LSCK-----AEGLEBESSIESRTL-----SWSYSMDOGR--	425
Qy	1000	LKPLYEQGALIKOLDANNVRHNAPODIOSKLETLDLGEHGAELLNDMKRFDELFOESA	1059
Db	426	-----NLIGSSDLFVIDEAGMVGSROLARFIGEAERG	458
Qy	1060	TRSVTVLGOHQGVLYKNGEINSEFKPSGKAL-----VQSFNVNSGGDLKSLOQAVH	1113
Db	459	AK-IVLVGDHE-----QLQALGACAPPRAIAEQTHVELSGSIRRQRHDWQQAQSA-F	509
Qy	1114	ATPPSAESKLQSMGLHFVSAGVDMSHOKGEIPLGRORDPNDKTALTKSLILDVTVTIGEL	1173
Db	510	AFTKTAE-----GLAAVRDHDGIHFAESRD-----	534
Qy	1174	HELADAKALKVSDHKPDADQIKOLROOFDTLRKRRYESNPVKHYTMGFTHNKALEANYDA	1233
Db	535	---AAMAQIVRDYIDD-----NEKRDPG-----TRYAMAHRR---ADVRA	568
Qy	1234	VKAFINAPKKEHH-----GYNLTRTVLESQGS---AELAKLKNLTLLSIDSGE	1279
Db	569	LNATIRSELQNRQLERSLGLSDGPDGRGDVEDRGNSGDVLAELTFQTSNCKRAFASGD	628
Qy	1280	SMSF-----SHSYCGGVSTVPVPTLSKKV-----PVPI	1308
Db	629	RIFLENNRDLGVKNMGMLCTVEDVEKGRIVARLDRGDSGSIPTDSYQAI DHGYATTIH	688
Qy	1309	PGAGITLDRAYNLFSRTSGGLNVSFGRDGGVSGNIMVATGHDVPMYPYTGKK-TSAGNAS	1367
Db	689	KNOGATVDRAFVASST-----MDRHLAYVMTPRHRSVOLYADIKEFTSAGRVL	738
Qy	1368	DWLSAKHKISP-----DLRTGAASVGTLOGCTLQNSUKFLKLTDELDP--GF IH	1412
Db	739	D-----HGVAPYEHNRRARENIFYVTLENDKGERHTVMGVDLN-----RAMRDASAIGDKI	789
Qy	1413	GLTHGTLTPAEL-----LOKGIE-HOMQGSKLTFFSDVTSANLD-----RA	1453
Db	790	GLQHVGATPVTLPDGTKAERNANRWVKEELAYOKLASRJSRKAKETLTDYISEFAERR	849

Qy	1454	GINLNEDGSKPVGTVARVSAGLSASANAAGSRERSTTSGQFSTTSASNNRPT-----	1507
Db	850	GIASDPRIAGDOGIADQL--CLKSEIELASARKDQEVLSRAAHLOREQQDRPSEROOVY	907
Qy	1508	-----FLNGVYGAGANLTA--	1521
Db	908	EEERAGLAGVVRGRHPRSPFARDIDEGQDDRAEETNRYRIRWSDLSMQDGAVPPTET	967
Qy	1522	LGVAHSTTHEKGKVGVIFFAPFTSTNVSAALADNRTSOSI-SLELKRAEPTYNSDISELTS	1580
Db	968	VG-QQSSLENGRPTPLVPAIT-----RYQRSIEVAORALSVIDOQFDTVES	1014
Qy	1581	TICKHKFKDSA--TTKMLAALKELDDAKPABOLHILOOHFSKADVGD-ERYEAVR----	1632
Db	1015	LVRVFRDPAEVAARLFAAMTEKEG-----NGKINAKAMAGOPERFELRGESGL	1064
Qy	1633	--NLKKLVTROQAADSHSMELGSAHSSTTYNNLSRINNDOIVELLHKHFDAAALPASSAKR	1690
Db	1065	FGSNKERKEALQYARSLSAHIGVVSAAW-----ERR	1095
Qy	1691	LGEMMNNDP---ALKDIIKOLOSTPFFSASVSMELDKGLREOTEKAILDGKVGREEVGVL	1747
Db	1096	LGEEFROSEOWREQRDVIEVPGILTP-RSAEILAKVEMQVEKRGKIKD-----	1143
Qy	1748	FODRNNLRKVSVSQSVSKSGGFNTPALLLGTSNSAAMSMERNICTINFKYQGDQONTTPR	1807
Db	1144	-----LR-----SSAEG-----QAALDEARQVADALTRFRGS-----SDPR	1174
Qy	1808	RFTLEGGIAQANPOVA	1823
Db	1175	RFAGE--LEARPELA	1187

```

RESULT 14
PCT-US01-04098A-3647
; Sequence 3647, Application PC/TUS0104098A
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-029
; CURRENT APPLICATION NUMBER: PCT/US01/04098A
; CURRENT FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: Not Yet Assigned
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 09/728,422
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 09/693,325
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 09/663,561
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: 09/654,936
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 09/620,325
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/598,075
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 3960
; SOFTWARE: Custom
; SEQ ID NO 3647
; LENGTH: 1654
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US01-04098A-3647

```

Query Match 1.7%; Score 163.5; DB 1; Length 1654;  
Best Local Similarity 18.7%; Pred. NO. 0.32;  
Matches 322; Conservative 198; Mismatches 606; Indels 599; Gaps 82;

QY	17	AAH-NPVGHGVALQOQSSSSPQNAASAIAECKNRG-----KMPRIHQPS-----	61
Db	59	AAHPDPKSTSVDTGQVSRPNPQSPASPRVTCKARSPVRLPHGSPGPEKAAAPDDYS	118
QY	62	--TAADGISAHQKKSFSLRG-----LGTKKFSRPAQOPQPTTHSKGA	105
Db	119	KTRSASSTPHNTRRVAALRGAGPAGBAGMTWPAGAVLPDPLTSGQRQAGP-NISKAL	177
QY	106	TLRDOLLARD-----DGETQHEAAAPDAA-RLTRSGGVYKRRNMDMAGRPMVYGGSGE	156
Db	178	EMTGIHAPESQEPSSLLEGADSVSSRAPQASLSMLPSTDNTKEACGHVSGH-CCPGCSRE	236
QY	157	DKVPTOQKRHOJNFFGQMRQTMJLKMHAHPASANGD-----RLQHSPPHIGSHHEIKE	210
Db	237	SPV-----TDIDSFIKELDASAARSPSOTGDSQOEGSAQGHPPAGAGGGSCRA	287
QY	211	EPV--GSTSKATTA-----HA	224
Db	288	EPVPGGOTSPRRAWAAGAPAYPOWASQPSVLDINSIPDKHFTVKNKFLNSYRNFSSFHE	347
QY	225	DRVEIA-----QEDDDSEFOOLHQ-----QRLARERNP-253	
Db	348	DSTSLSLGLDSTEPSSLSSMYGDAEDSSDPSLTPAPRASARDGWSPSRVSLHKEDPS	407
QY	254	-----POPKLGVA TPISARQP-----KLTAV	276
Db	408	ESEBEQIETICTRGCPNPNPSPAHLPTQAAICPASAKVLSLYKSTPRESVASPREKVA CL	467
QY	277	ABSVLEGGTDTTQSLPKPOSMLKGSG-----AGVTP-----LAVTLDKGLQLAPDN	322
Db	468	PGSYTSGPDSQ-----PSSLLEMSQEHETHADISTQNHRRPSCAEETEVTSASSAMEN	523
QY	323	PPALNTLLKQTLGKDTQHY-----LAHHASSDGOHLLDNKGLHFDIKTATSYSVLH	376
Db	524	SP-----LSKVARHFHSPPIITLSPNMVNGLEHLLDDE-----TLNQYETSI	566
QY	377	NSHPGEIKLQAAGTGVSVDPKGSKTSLGSGTOSHNTKMTLSQPEAHRSLLTQIWOHP	436
Db	567	N-----AAASLSFSVD-----VPKNGBSVLENIHISEQDLDDLQKP	605
QY	437	AGAA R-----PGESIRLHDDKIHLHPELGWGSAOKDTHSQLSRQADGKL	483
Db	606	KMIARRPIMAWFKEINKNQGTHLRSKTEKEQPLMPA-----RSPGSKTQWSSSKGVT	661
QY	484	Y-----ALKDNRTLONLSNKKSEKLV-----DKIKSYSDVQORGQVALTDTPGR-	528
Db	662	VPHSPPOPTKNLENKDLSSKSPAEMLLTNCQAKCGPKLRLSL--KGKAKVNSEAPAA N	719
QY	529	-----HKMSTMPSLDASPE-SHTLSL-----LHPAD-----AHQGLLH	560
Db	720	AVKAGGTDHRKPLI-----SPQTSHTLSKAVSQRLHVADHEDDDRTTAA PRSPOCVLE	774
QY	561	GKSELEAQSAVISHGRVLVADSEGLFSA-----AIPKQDGNELKMKAMPOHALDEHF	614
Db	775	SKBPATSGBP-----KPSVSDTSIRTFVSPLTSPKVPPEQMGWSRPHMVLSE-----	823
QY	615	GHDHQISGFFPHDDHQJUNALVKNFNROQHACPLGNHDHFHPGWNLLDALYIDN-----	669
Db	824	-----PDRCGCTTPKSPKCAFGRAPRADSGFVSPAASRNGMSVAGNRQSEPR	872
QY	670	GLH-----HTNPPEHILDMHGLSLAQEGKLHYFDOL-----TKWGTAESDCKQLKGL	721
Db	873	ASHVAALTAQPRP-----TGEKGNINIASDRLEFNTQKIVEISAEAVSETVCGNKPAS	927
QY	722	DGAAYLLKDGVEK-----RLMINQSTSIKHGTENFVLSPHVRNKPE-----	763
Db	928	DRRGCLAQGNCOEKSEIRLYRQVAESSTSHPS-----SLPASHASQAQBSMRSFMAKLA	983
QY	764	-PGDALQ-GLNKDDKAQAMAVI-----GV--NKYLALTEKGDIRSFOIKPCTQOLERP	814
Db	984	SSSSSLQTAIRKAEYSQKSSLSMDSRGVPRNSTIPGGSGEDHLYFTPRPATRYSMPAQ	1043

[illegible]

Query Match 1.7%; Score 163; DB 5; Length 1242;  
Best Local Similarity 18.8%; Pred.No. 0.2; Indels 530; Gaps  
Matches 256; Conservative 146; Mismatches 431;

Qy 282 EGDITTSPLKPSMLKGSGAGVTPLAVTL-----DKGILQLAPDNPPLATLLLKOTLKG 336



